Molecular Modeling and Dynamics

Modeling Metods, Molecular Dynamics, Computational Models and Computer Simulations of Molecular Systems

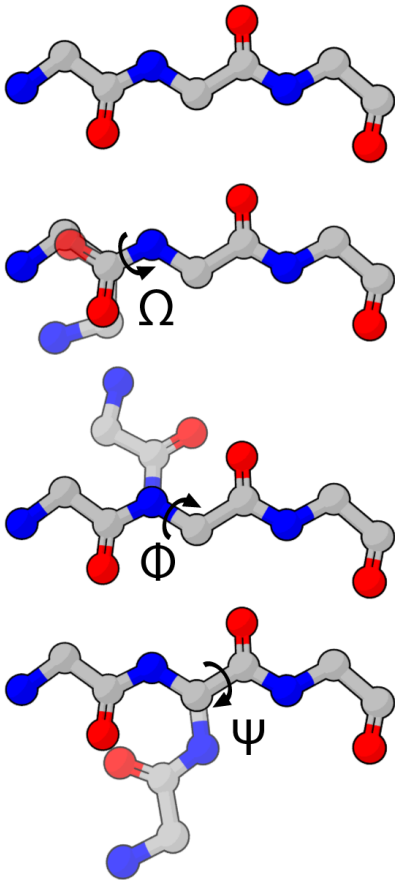
Molecular Modeling Metods

Molecular modeling

1. REDIRECT Molecular modelling

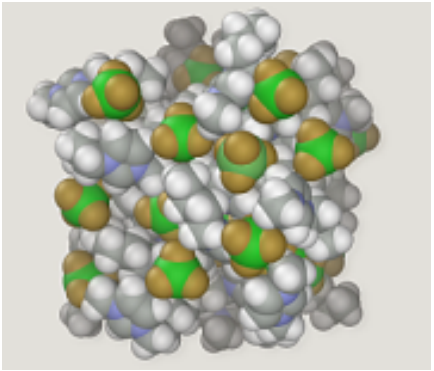
Molecular modelling

Molecular modelling is a collective term that refers to theoretical methods and computational techniques to model or mimic the behaviour of molecules. The techniques are used in the fields of computational chemistry, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modelling of any reasonably sized system. The common feature of molecular modelling techniques is the atomistic level description of the molecular systems; the lowest level of information is individual atoms (or a small group of atoms). This is in contrast to quantum chemistry (also known as electronic structure calculations) where electrons are considered explicitly. The benefit of molecular modelling is that it reduces the complexity of the system, allowing many more particles (atoms) to be considered during simulations.



The backbone dihedral angles are included in the molecular model of a protein.

Molecular mechanics is one aspect of molecular modelling, as it is refers to the use of classical mechanics/Newtonian mechanics to describe the physical basis behind the models. Molecular models typically describe atoms (nucleus and electrons collectively) as point charges with an associated mass. The interactions between neighbouring atoms are described by spring-like interactions (representing chemical bonds) and van der Waals forces. The Lennard-Jones potential is commonly used to describe van der Waals forces. The electrostatic interactions are computed based on Coulomb's law. Atoms are assigned



Modelling of ionic liquid

coordinates in Cartesian space or in internal coordinates, and can also be assigned velocities in dynamical simulations. The atomic velocities are related to the temperature of the system, a macroscopic quantity. The collective mathematical expression is known as a potential function and is related to the system internal energy (U), a thermodynamic quantity equal to the sum of potential and kinetic energies. Methods which minimize the potential energy are known as energy minimization techniques (e.g., steepest descent and conjugate gradient), while methods that model the behaviour of the system with propagation of time are known as molecular dynamics.

$$E = E_{bonds} + E_{angle} + E_{dihedral} + E_{non-bonded}$$

$$E_{non-bonded} = E_{electrostatic} + E_{vanderWaals}$$

This function, referred to as a potential function, computes the molecular potential energy as a sum of energy terms that describe the deviation of bond lengths, bond angles and torsion angles away from equilibrium values, plus terms for non-bonded pairs of atoms describing van der Waals and electrostatic interactions. The set of parameters consisting of equilibrium bond lengths, bond angles, partial charge values, force constants and van der Waals parameters are collectively known as a force field. Different implementations of molecular mechanics use slightly different mathematical expressions, and therefore, different constants for the potential function. The common force fields in use today have been developed by using high level quantum calculations and/or fitting to experimental data. The technique known as energy minimization is used to find positions of zero gradient for all atoms, in other words, a local energy minimum. Lower energy states are more stable and are commonly investigated because of their role in chemical and biological processes. A molecular dynamics simulation, on the other hand, computes the behaviour of a system as a function of time. It involves solving Newton's laws of motion, principally the second law, **F = ma**. Integration of Newton's laws of motion, using different integration algorithms, leads to atomic trajectories in space and time. The force on an atom is defined as the negative gradient of the potential energy function. The energy minimization technique is useful for obtaining a static picture for comparing between states of similar systems, while molecular dynamics provides information about the dynamic processes with the intrinsic inclusion of temperature effects.

Molecules can be modelled either in vacuum or in the presence of a solvent such as water. Simulations of systems in vacuum are referred to as *gas-phase* simulations, while those that include the presence of solvent molecules are referred to as *explicit solvent* simulations. In another type of simulation, the effect of solvent is estimated using an empirical mathematical expression; these are known as *implicit solvation* simulations.

Molecular modelling methods are now routinely used to investigate the structure, dynamics and thermodynamics of inorganic, biological, and polymeric systems. The types of biological activity that have been investigated using molecular modelling include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes.

Popular software for molecular modelling

- Abalone
- AMBER
- ADF
- Ascalaph Designer[1]
- BALLView
- Biskit
- BOSS
- Cerius2
- Chimera
- CHARMM
- Coot (program)[2] for X-ray crystallography of biological molecules
- COSMOS (software)[3]
- CP2K
- CPMD
- Firefly
- GAMESS (UK)
- GAMESS (US)
- GAUSSIAN
- Ghemical
- GROMACS
- GROMOS
- InsightII
- LAMMPS
- MacroModel
- MarvinSpace[4]
- Materials Studio
- MDynaMix
- MMTK
- MOE (software)[5]
- Molecular Docking Server
- Molsoft ICM[6]
- MOPAC
- NAMD
- NOCH
- Oscail X
- PyMOL
- Q-Chem
- Sirius
- SPARTAN (software)[7]
- STR3DI32[8]
- Sybyl (software)[9]
- MCCCS Towhee[10]
- TURBOMOLE
- ReaxFF
- VMD
- WHAT IF[11]

- xeo[12]
- YASARA[13]
- Zodiac (software)[14]

See also

- Cheminformatics
- Computational chemistry
- Density functional theory programs.
- Force field in Chemistry
- Force field implementation
- List of nucleic acid simulation software
- List of protein structure prediction software
- Molecular Design software
- Molecular dynamics
- Molecular graphics
- Molecular mechanics
- Molecular model
- Molecular modelling on GPU
- Molecule editor
- Monte Carlo method
- Quantum chemistry computer programs
- Semi-empirical quantum chemistry method
- Software for molecular mechanics modelling
- Structural Bioinformatics

External links

- Center for Molecular Modeling at the National Institutes of Health (NIH) [15] (U.S. Government Agency)
- Molecular Simulation [16], details for the Molecular Simulation journal, ISSN: 0892-7022 (print), 1029-0435 (online)
- The eCheminfo [17] Network and Community of Practice in Informatics and Modeling

References

- M. P. Allen, D. J. Tildesley, *Computer simulation of liquids*, 1989, Oxford University Press, ISBN 0-19-855645-4.
- A. R. Leach, *Molecular Modelling: Principles and Applications*, 2001, ISBN 0-582-38210-6
- D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 1996, ISBN 0-12-267370-0
- D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, 2004, ISBN 0-521-82586-7
- R. J. Sadus, *Molecular Simulation of Fluids: Theory, Algorithms and Object-Orientation*, 2002, ISBN 0-444-51082-6
- K.I.Ramachandran, G Deepa and Krishnan Namboori. P.K. *Computational Chemistry and Molecular Modeling Principles and Applications* 2008[18] ISBN 978-3-540-77302-3 Springer-Verlag GmbH

Homepage

[1] *Agile Molecule* (http://www.agilemolecule.com/index.html)
[2] *York Structural Biology Laboratory* (http://www.ysbl.york.ac.uk/~emsley/coot/)
[3] COSMOS (http://www.cosmos-software.de/ce_intro.html) - **C**omputer **S**imulation of **Mo**lecular **S**tructures
[4] *ChemAxon* (http://www.chemaxon.com/product/mspace.html)
[5] MOE - **M**olecular **O**perating **E**nvironment, *Chemical Computing Group* (http://www.chemcomp.com/)
[6] *Molsoft* (http://www.molsoft.com/)
[7] *Wavefunction, Inc.* (http://www.wavefun.com/)
[8] *Exorga, Inc.* (http://www.exorga.com/)
[9] *Tipos* (http://www.tripos.com/sybyl/)
[10] *MCCCS Towhee* (http://towhee.sourceforge.net/) - **M**onte **C**arlo for **C**omplex **C**hemical **S**ystems
[11] *CMBI* (http://swift.cmbi.ru.nl/whatif/)
[12] *xeo* (http://sourceforge.net/projects/xeo)
[13] *YASARA* (http://www.yasara.org/)
[14] *ZedeN* (http://www.zeden.org)
[15] http://cmm.info.nih.gov/modeling/
[16] http://www.tandf.co.uk/journals/titles/08927022.asp
[17] http://www.echeminfo.com/
[18] http://www.amrita.edu/cen/ccmm

Quantum chemistry

Quantum chemistry is a branch of theoretical chemistry, which applies quantum mechanics and quantum field theory to address issues and problems in chemistry. The description of the electronic behavior of atoms and molecules as pertaining to their reactivity is one of the applications of quantum chemistry. Quantum chemistry lies on the border between chemistry and physics, and significant contributions have been made by scientists from both fields. It has a strong and active overlap with the field of atomic physics and molecular physics, as well as physical chemistry.

Quantum chemistry mathematically describes the fundamental behavior of matter at the molecular scale.[1] It is, in principle, possible to describe all chemical systems using this theory. In practice, only the simplest chemical systems may realistically be investigated in purely quantum mechanical terms, and approximations must be made for most practical purposes (e.g., Hartree-Fock, post Hartree-Fock or Density functional theory, see computational chemistry for more details). Hence a detailed understanding of quantum mechanics is not necessary for most chemistry, as the important implications of the theory (principally the orbital approximation) can be understood and applied in simpler terms.

In quantum mechanics the Hamiltonian, or the physical state, of a particle can be expressed as the sum of two operators, one corresponding to kinetic energy and the other to potential energy. The Hamiltonian in the Schrödinger wave equation used in quantum chemistry does not contain terms for the spin of the electron.

Solutions of the Schrödinger equation for the hydrogen atom gives the form of the wave function for atomic orbitals, and the relative energy of the various orbitals. The orbital approximation can be used to understand the other atoms e.g. helium, lithium and carbon.

History

The **history of quantum chemistry** essentially began with the 1838 discovery of cathode rays by Michael Faraday, the 1859 statement of the black body radiation problem by Gustav Kirchhoff, the 1877 suggestion by Ludwig Boltzmann that the energy states of a physical system could be discrete, and the 1900 quantum hypothesis by Max Planck that any energy radiating atomic system can theoretically be divided into a number of discrete energy elements ε such that each of these energy elements is proportional to the frequency ν with which they each individually radiate energy, as defined by the following formula:

$$\epsilon = h\nu$$

where h is a numerical value called Planck's Constant. Then, in 1905, to explain the photoelectric effect (1839), i.e., that shining light on certain materials can function to eject electrons from the material, Albert Einstein postulated, based on Planck's quantum hypothesis, that light itself consists of individual quantum particles, which later came to be called photons (1926). In the years to follow, this theoretical basis slowly began to be applied to chemical structure, reactivity, and bonding.

Electronic structure

The first step in solving a quantum chemical problem is usually solving the Schrödinger equation (or Dirac equation in relativistic quantum chemistry) with the electronic molecular Hamiltonian. This is called determining the **electronic structure** of the molecule. It can be said that the electronic structure of a molecule or crystal implies essentially its chemical properties.

Wave model

The foundation of quantum mechanics and quantum chemistry is the **wave model**, in which the atom is a small, dense, positively charged nucleus surrounded by electrons. Unlike the earlier Bohr model of the atom, however, the wave model describes electrons as "clouds" moving in orbitals, and their positions are represented by probability distributions rather than discrete points. The strength of this model lies in its predictive power. Specifically, it predicts the pattern of chemically similar elements found in the periodic table. The wave model is so named because electrons exhibit properties (such as interference) traditionally associated with waves. See wave-particle duality.

Valence bond

Although the mathematical basis of quantum chemistry had been laid by Schrödinger in 1926, it is generally accepted that the first true calculation in quantum chemistry was that of the German physicists Walter Heitler and Fritz London on the hydrogen (H_2) molecule in 1927. Heitler and London's method was extended by the American theoretical physicist John C. Slater and the American theoretical chemist Linus Pauling to become the **Valence-Bond (VB)** [or **Heitler-London-Slater-Pauling (HLSP)**] method. In this method, attention is primarily devoted to the pairwise interactions between atoms, and this method therefore correlates closely with classical chemists' drawings of bonds.

Molecular orbital

An alternative approach was developed in 1929 by Friedrich Hund and Robert S. Mulliken, in which electrons are described by mathematical functions delocalized over an entire molecule. The **Hund-Mulliken** approach or **molecular orbital (MO) method** is less intuitive to chemists, but has turned out capable of predicting spectroscopic properties better than the VB method. This approach is the conceptional basis of the **Hartree-Fock method** and further post Hartree-Fock methods.

Density functional theory

The **Thomas-Fermi model** was developed independently by Thomas and Fermi in 1927. This was the first attempt to describe many-electron systems on the basis of electronic density instead of wave functions, although it was not very successful in the treatment of entire molecules. The method did provide the basis for what is now known as **density functional theory**. Though this method is less developed than post Hartree-Fock methods, its lower computational requirements allow it to tackle larger polyatomic molecules and even macromolecules, which has made it the most used method in computational chemistry at present.

Chemical dynamics

A further step can consist of solving the Schrödinger equation with the total molecular Hamiltonian in order to study the motion of molecules. Direct solution of the Schrödinger equation is called *quantum molecular dynamics*, within the semiclassical approximation *semiclassical molecular dynamics*, and within the classical mechanics framework *molecular dynamics (MD)*. Statistical approaches, using for example Monte Carlo methods, are also possible.

Adiabatic chemical dynamics

Main article: Adiabatic formalism or Born-Oppenheimer approximation

In **adiabatic dynamics**, interatomic interactions are represented by single scalar potentials called potential energy surfaces. This is the Born-Oppenheimer approximation introduced by Born and Oppenheimer in 1927. Pioneering applications of this in chemistry were performed by Rice and Ramsperger in 1927 and Kassel in 1928, and generalized into the RRKM theory in 1952 by Marcus who took the transition state theory developed by Eyring in 1935 into account. These methods enable simple estimates of unimolecular reaction rates from a few characteristics of the potential surface.

Non-adiabatic chemical dynamics

Non-adiabatic dynamics consists of taking the interaction between several coupled potential energy surface (corresponding to different electronic quantum states of the molecule). The coupling terms are called **vibronic couplings**. The pioneering work in this field was done by Stueckelberg, Landau, and Zener in the 1930s, in their work on what is now known as the Landau-Zener transition. Their formula allows the transition probability between two diabatic potential curves in the neighborhood of an avoided crossing to be calculated.

Quantum chemistry and quantum field theory

The application of quantum field theory (QFT) to chemical systems and theories has become increasingly common in the modern physical sciences. One of the first and most fundamentally explicit appearances of this is seen in the theory of the photomagneton. In this system, plasmas, which are ubiquitous in both physics and chemistry, are studied in order to determine the basic quantization of the underlying bosonic field. However, quantum field theory is of interest in many fields of chemistry, including: nuclear chemistry, astrochemistry, sonochemistry, and quantum hydrodynamics. Field theoretic methods have also been critical in developing the ab initio Effective Hamiltonian theory of semi-empirical pi-electron methods.

See also

- Atomic physics
- Computational chemistry
- Condensed matter physics
- International Academy of Quantum Molecular Science
- Physical chemistry
- Quantum chemistry computer programs
- Quantum electrochemistry
- QMC@Home
- Theoretical physics

Further reading

- Pauling, L. (1954). *General Chemistry*. Dover Publications. ISBN 0-486-65622-5.
- Pauling, L., and Wilson, E. B. *Introduction to Quantum Mechanics with Applications to Chemistry* (Dover Publications) ISBN 0-486-64871-0
- Atkins, P.W. *Physical Chemistry* (Oxford University Press) ISBN 0-19-879285-9
- McWeeny, R. *Coulson's Valence* (Oxford Science Publications) ISBN 0-19-855144-4
- Landau, L.D. and Lifshitz, E.M. *Quantum Mechanics:Non-relativistic Theory* (Course of Theoretical Physics vol.3) (Pergamon Press)
- Bernard Pullman and Alberte Pullman. 1963. *Quantum Biochemistry.*, New York and London: Academic Press.
- Eric R. Scerri, The Periodic Table: Its Story and Its Significance, Oxford University Press, 2006. Considers the extent to which chemistry and especially the periodic system has been reduced to quantum mechanics. ISBN 0-19-530573-6.
- Simon, Z. 1976. *Quantum Biochemistry and Specific Interactions.*, Taylor & Francis; ISBN-13: 978-0856260872 and ISBN 0-85-6260878 .

References

[1] http://cmm.cit.nih.gov/modeling/guide_documents/quantum_mechanics_document.html|"Quantum Chemistry". *The NIH Guide to Molecular Modeling*. National Institutes of Health. http://cmm.cit.nih.gov/modeling/guide_documents/quantum_mechanics_document.html. Retrieved on 2007-09-08.

External links

- The Sherrill Group - Notes (http://vergil.chemistry.gatech.edu/notes/index.html)
- ChemViz Curriculum Support Resources (http://www.shodor.org/chemviz/)
- Early ideas in the history of quantum chemistry (http://www.quantum-chemistry-history.com/)

Nobel lectures by quantum chemists

- Walter Kohn's Nobel lecture (http://nobelprize.org/chemistry/laureates/1998/kohn-lecture.html)
- Rudolph Marcus' Nobel lecture (http://nobelprize.org/chemistry/laureates/1992/marcus-lecture.html)
- Robert Mulliken's Nobel lecture (http://nobelprize.org/chemistry/laureates/1966/mulliken-lecture.html)
- Linus Pauling's Nobel lecture (http://nobelprize.org/chemistry/laureates/1954/pauling-lecture.html)
- John Pople's Nobel lecture (http://nobelprize.org/chemistry/laureates/1998/pople-lecture.html)

Molecular orbital theory

In chemistry, **molecular orbital theory** (*MO theory*) is a method for determining molecular structure in which electrons are not assigned to individual bonds between atoms, but are treated as moving under the influence of the nuclei in the whole molecule.[1] In this theory, each molecule has a set of molecular orbitals, in which it is assumed that the molecular orbital wave function ψ_f may be written as a simple weighted sum of the n constituent atomic orbitals χ_i, according to the following equation:[2]

$$\psi_j = \sum_{i=1}^{n} c_{ij}\chi_i$$

The c_{ij} coefficients may be determined numerically by substitution of this equation into the Schrödinger equation and application of the variational principle. This method is called the linear combination of atomic orbitals approximation and is used in computational chemistry. An additional unitary transformation can be applied on the system to accelerate the convergence in some computational schemes. Molecular orbital theory was seen as a competitor to valence bond theory in the 1930s, before it was realized that the two methods are closely related and that when extended they become equivalent.

History

Molecular orbital theory was developed, in the years after valence bond theory (1927) had been established, primarily through the efforts of Friedrich Hund, Robert Mulliken, John C. Slater, and John Lennard-Jones.[3] MO theory was originally called the Hund-Mulliken theory.[4] The word *orbital* was introduced by Mulliken in 1932.[4] By 1933, the molecular orbital theory had become accepted as a valid and useful theory.[5] According to German physicist and physical chemist Erich Hückel, the first quantitative use of molecular orbital theory was the 1929 paper of Lennard-Jones.[6] The first accurate calculation of a molecular orbital wavefunction was that made by Charles Coulson in 1938 on the hydrogen molecule.[7] By 1950, molecular orbitals were completely defined as eigenfunctions (wave functions) of the self-consistent field Hamiltonian and it was at this point that molecular orbital theory became fully rigorous and consistent.[8] This rigorous approach is known as the Hartree-Fock method for molecules although it had its origins in calculations on atoms. In calculations on molecules, the molecular orbitals are expanded in terms of an atomic orbital basis set, leading to the Roothaan equations.[9] This led to the development of many ab initio quantum chemistry methods. Parallel to this rigorous development, molecular orbital theory was applied in an approximate manner using some empirically derived parameters in methods now known as semi-empirical quantum chemistry methods.[10]

Overview

Molecular orbital (MO) theory uses a linear combination of atomic orbitals to form molecular orbitals which cover the whole molecule. These are often divided into bonding orbitals, anti-bonding orbitals, and non-bonding orbitals. A molecular orbital is merely a Schrödinger orbital which includes several, but often only two nuclei. If this orbital is of type in which the electron(s) in the orbital have a higher probability of being *between* nuclei than elsewhere, the orbital will be a bonding orbital, and will tend to hold the nuclei together. If the electrons tend to be present in a molecular orbital in which they spend more time elsewhere than between the nuclei, the orbital will function as an anti-bonding orbital and will actually weaken the bond. Electrons in non-bonding orbitals tend to be in deep orbitals (nearly atomic orbitals) associated almost entirely with one nucleus or the other, and thus they spend equal time between nuclei or not. These electrons neither contribute nor detract from bond strength.

Molecular orbitals are further divided according to the types of atomic orbitals combining to form a bond. These orbitals are results of electron-nucleus interactions that are caused by the fundamental force of electromagnetism. Chemical substances will form a bond if their orbitals become lower in energy when they interact with each other. Different chemical bonds are distinguished that differ by electron cloud shape and by energy levels.

MO theory provides a global, delocalized perspective on chemical bonding. For example, in the MO theory for hypervalent molecules it is unnecessary to invoke a major role for d-orbitals, whereas valence bond theory normally uses hybridization with d-orbitals to explain hypervalency. In MO theory, *any* electron in a molecule may be found *anywhere* in the molecule, since quantum conditions allow electrons to travel under the influence of an arbitrarily large number of nuclei, so long as permitted by certain quantum rules. Although in MO theory *some* molecular orbitals may hold electrons which are more localized between specific pairs of molecular atoms, *other* orbitals may hold electrons which are spread more uniformly over the molecule. Thus, overall, bonding (and electrons) are far more delocalized

(spread out) in MO theory, than is implied in valence bond (VB) theory. This makes MO theory more useful for the description of extended systems.

An example is that in the MO picture of benzene, composed of a hexagonal ring of 6 carbon atoms. In this molecule, 24 of the 30 total valence bonding electrons are located in 12 σ (sigma) bonding orbitals which are mostly located between pairs of atoms (C-C or C-H), similar to the valence bond picture. However, in benzene the remaining 6 bonding electrons are located in 3 π (pi) molecular bonding orbitals that are delocalized around the ring. Two are in an MO which has equal contributions from all 6 atoms. The other two orbitals have vertical nodes at right angles to each other. As in the VB theory, all of these 6 delocalized pi electrons reside in a larger space which exists above and below the ring plane. All carbon-carbon bonds in benzene are chemically equivalent. In MO theory this is a direct consequence of the fact that the 3 molecular pi orbitals form a combination which evenly spreads the extra 6 electrons over 6 carbon atoms.[11]

In molecules such as methane, the 8 valence electrons are found in 4 MOs that are spread out over all 5 atoms. However, it is possible to approximate the MOs with 4 localized orbitals similar in shape to sp^3 hybrid orbitals predicted by VB theory. This is often adequate for σ (sigma) bonds, but it is not possible for the π (pi) orbitals. However, the delocalized MO picture is more appropriate for ionization and spectroscopic predictions. Upon ionization of methane, a single electron is taken from the MO which surrounds the whole molecule, weakening all 4 bonds equally. VB theory would predict that one electron is removed for an sp^3 orbital, resulting in the need for resonance between four valence bond structures, each of which has a one-electron bond.

As in benzene, in substances such as beta carotene, chlorophyll or heme, some electrons the π (pi) orbitals are spread out in molecular orbitals over long distances in a molecule, giving rise to light absorption in lower energies (visible colors), a fact which is observed. This and other spectroscopic data for molecules are better explained in MO theory, with an emphasis on electronic states associated with multicenter orbitals, including mixing of orbitals premised on principles of orbital symmetry matching. The same MO principles also more naturally explain some electrical phenomena, such as high electrical conductivity in the planar direction of the hexagonal atomic sheets that exist in graphite. In MO theory, "resonance" (a mixing and blending of VB bond states) is a natural consequence of symmetry. For example, in graphite, as in benzene, it is not necessary to invoke the sp^2 hybridization and resonance of VB theory, in order to explain electrical conduction. Instead, MO theory simply recognizes that some electrons in the graphite atomic sheets are completely delocalized over arbitrary distances, and reside in very large *molecular orbitals* that cover an entire graphite sheet, and some electrons are thus as free to move and conduct electricity *in the sheet plane,* as if they resided in a metal.

See also

- Ab initio quantum chemistry methods
- Atomic orbital
- Configuration interaction
- Coupled cluster
- Hartree-Fock
- Molecular orbital
- MO diagram
- Møller-Plesset perturbation theory
- Quantum chemistry computer programs
- Semi-empirical quantum chemistry methods

References

[1] Daintith, J. (2004). *Oxford Dictionary of Chemistry*. New York: Oxford University Press. ISBN 0-19-860918-3.
[2] Licker, Mark, J. (2004). *McGraw-Hill Concise Encyclopedia of Chemistry*. New York: McGraw-Hill. ISBN 0-07-143953-6.
[3] Coulson, Charles, A. (1952). *Valence*. Oxford at the Clarendon Press.
[4] Spectroscopy, Molecular Orbitals, and Chemical Bonding (http://nobelprize.org/nobel_prizes/chemistry/laureates/1966/mulliken-lecture.pdf) - Robert Mulliken's 1966 Nobel Lecture
[5] Lennard-Jones Paper of 1929 (http://www.quantum-chemistry-history.com/LeJo_Dat/LJ-Hall1.htm) - Foundations of Molecular Orbital Theory.
[6] Hückel, E. (1934). *Trans. Faraday Soc. 30*, 59.
[7] Coulson, C.A. (1938). *Proc. Camb. Phil. Soc. 34*, 204.
[8] Hall, G.G. Lennard-Jones, Sir John. (1950). *Proc. Roy. Soc. A202*, 155.
[9] Frank Jensen, Introduction to Computational Chemistry, John Wiley and Sons, 1999, pg 65 - 69, ISBN 0 471 98055
[10] Frank Jensen, Introduction to Computational Chemistry, John Wiley and Sons, 1999, pg 81 - 92, ISBN 0 471 98055
[11] Introduction to Molecular Orbital Theory (http://www.ch.ic.ac.uk/vchemlib/course/mo_theory/main.html) - Imperial College London

External links

- Molecular Orbital Theory (http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch8/mo.html) - Purdue University
- Molecular Orbital Theory (http://www.sparknotes.com/chemistry/bonding/molecularorbital/section1.html) - Sparknotes
- Molecular Orbital Theory (http://www.mpcfaculty.net/mark_bishop/molecular_orbital_theory.htm) - Mark Bishop's Chemistry Site
- Introduction to MO Theory (http://www.chem.qmul.ac.uk/software/download/mo/) - Queen Mary, London University
- Molecular Orbital Theory (http://www.chm.davidson.edu/ChemistryApplets/MolecularOrbitals/index.html) - a related terms table

Linear combination of atomic orbitals molecular orbital method

Electronic structure methods
Tight binding
Nearly-free electron model
Hartree–Fock
Modern valence bond
Generalized valence bond
Møller–Plesset perturbation theory
Configuration interaction
Coupled cluster
Multi-configurational self-consistent field
Density functional theory
Quantum chemistry composite methods
Quantum Monte Carlo
k·p perturbation theory
Muffin-tin approximation
LCAO method

A **linear combination of atomic orbitals** or **LCAO** is a quantum superposition of atomic orbitals and a technique for calculating molecular orbitals in quantum chemistry.[1] In quantum mechanics, electron configurations of atoms are described as wavefunctions. In mathematical sense, these wave functions are the basis set of functions, the basis functions, which describe the electrons of a given atom. In chemical reactions, orbital wavefunctions are modified, i.e. the electron cloud shape is changed, according to the type of atoms participating in the chemical bond.

It was introduced in 1929 by Sir John Lennard-Jones with the description of bonding in the diatomic molecules of the first main row of the periodic table, but had been used earlier by Linus Pauling for H_2^+.[2] [3]

A mathematical description is

$$\phi_i = c_1\chi_1 + c_2\chi_2 + c_3\chi_3 + \cdots + c_n\chi_n$$

or

$$\phi_i = \sum_r c_{ri}\chi_r$$

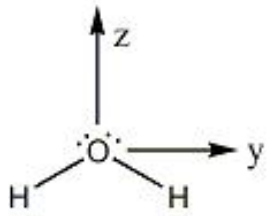
where ϕ_i (phi) is a molecular orbital represented as the sum of n atomic orbitals χ_r (chi), each multiplied by a corresponding coefficient c_r. The coefficients are the weights of the contributions of the n atomic orbitals to the molecular orbital. The Hartree-Fock procedure is used to obtain the coefficients of the expansion from the Hartree-Fock procedure.

The orbitals are thus expressed as linear combinations of basis functions, and the basis functions are one-electron functions centered on nuclei of the component atoms of the molecule. The atomic orbitals used are typically those of hydrogen-like atoms since these

are known analytically i.e. Slater-type orbitals but other choices are possible like Gaussian functions from standard basis sets.

By minimizing the total energy of the system, an appropriate set of coefficients of the linear combinations is determined. This quantitative approach is now known as the Hartree-Fock method. However, since the development of computational chemistry, the LCAO method often refers not to an actual optimization of the wave function but to a qualitative discussion which is very useful for predicting and rationalizing results obtained via more modern methods. In this case, the shape of the molecular orbitals and their respective energies are deduced approximately from comparing the energies of the atomic orbitals of the individual atoms (or molecular fragments) and applying some recipes known as level repulsion and the like. The graphs that are plotted to make this discussion clearer are called **correlation diagrams**. The required atomic orbital energies can come from calculations or directly from experiment via Koopmans' theorem.

This is done by using the symmetry of the molecules and orbitals involved in bonding. The first step in this process is assigning a point group to the molecule. A common example is water, which is of C_{2v} symmetry. Then a reducible representation of the bonding is determined demonstrated below for water:

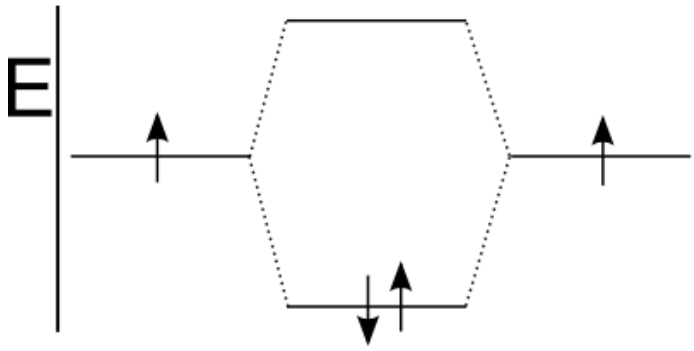


C_{2v}	E	C2	$\sigma_v(xz)$	$\sigma_v'(yz)$
$\Gamma\sigma$	2	0	0	2

$$\Gamma\sigma = A_1 + B_2$$

Each operation in the point group is performed upon the molecule. The number of bonds that are unmoved is the character of that operation. This reducible representation is decomposed into the sum of irreducible representations. These irreducible representations correspond to the symmetry of the orbitals involved.

MO diagrams provide simple qualitative LCAO treatment.



Quantitative theories are the Huckel method, the extended Huckel method and the Pariser–Parr–Pople method.

See also

- Quantum chemistry computer programs
- Hartree-Fock
- Basis set (chemistry)
- Tight binding

External links

- LCAO @ chemistry.umeche.maine.edu Link [4]

References

[1] Huheey, James. *Inorganic Chemistry:Principles of Structure and Reactivity*
[2] Friedrich Hund and Chemistry, Werner Kutzelnigg, on the occasion of Hund's 100th birthday, Angewandte Chemie, 35, 573 - 586, (1996)
[3] Robert S. Mulliken's Nobel Lecture, Science, 157, no. 3785, 13 - 24, (1967)
[4] http://chemistry.umeche.maine.edu/Modeling/lcao.html

Hückel method

The **Hückel method** or **Hückel molecular orbital method** (HMO) proposed by Erich Hückel in 1930, is a very simple linear combination of atomic orbitals molecular orbitals (LCAO MO) method for the determination of energies of molecular orbitals of pi electrons in conjugated hydrocarbon systems, such as ethene, benzene and butadiene. [1] [2] It is the theoretical basis for the Hückel's rule; the extended Hückel method developed by Roald Hoffmann is the basis of the Woodward-Hoffmann rules [3] . It was later extended to conjugated molecules such as pyridine, pyrrole and furan that contain atoms other than carbon, known in this context as heteroatoms. [4]

It is a very powerful educational tool and details appear in many chemistry textbooks.

Hückel characteristics

The method has several characteristics:

- It limits itself to conjugated hydrocarbons
- Only pi electron MO's are included because these determine the general properties of these molecules and the sigma electrons are ignored. This is referred to as sigma-pi separability.
- The method takes as inputs the LCAO MO Method, the Schrödinger equation and simplifications based on orbital symmetry considerations. Interestingly the method does not take in any physical constants.
- The method predicts how many energy levels exist for a given molecule, which levels are degenerate and it expresses the MO energies as the sum of two other energy terms called alpha, the energy of an electron in a 2p-orbital and beta, an interaction energy between two p orbitals which are still unknown but importantly have become independent of the molecule. In addition it enables calculation of charge density for each atom in the pi framework, the bond order between any two atoms and the overall molecular dipole moment.

Hückel results

The results for a few simple molecules are tabulated below:

Molecule	Energy	Frontier orbital	HOMO - LUMO energy gap
Ethylene	$E_1 = \alpha - \beta$	LUMO	-2β
	$E_2 = \alpha + \beta$	HOMO	
Butadiene	$E_1 = \alpha + 1.62\beta$		
	$E_2 = \alpha + 0.62\beta$	HOMO	-1.24β
	$E_3 = \alpha - 0.62\beta$	LUMO	
	$E_4 = \alpha - 1.62\beta$		
Benzene	$E_1 = \alpha + 2\beta$		
	$E_2 = \alpha + \beta$		
	$E_3 = \alpha + \beta$	HOMO	-2β
	$E_4 = \alpha - \beta$	LUMO	
	$E_5 = \alpha - \beta$		
	$E_6 = \alpha - 2\beta$		
Cyclobutadiene	$E_1 = \alpha + 2\beta$		
	$E_2 = \alpha$	SOMO	0
	$E_3 = \alpha$	SOMO	
	$E_4 = \alpha - 2\beta$		

Table 1. Hückel method results Lowest energies op top α and β are both negative values [5]

The theory predicts two energy levels for ethylene with its two pi electrons filling the low-energy HOMO and the high energy LUMO remaining empty. In butadiene the 4 pi electrons occupy 2 low energy MO's out of a total of 4 and for benzene 6 energy levels are predicted two of them degenerate.

For linear and cyclic systems (with n atoms), general solutions exist [6] .

Linear: $E_k = \alpha + 2\beta \cos \frac{k\pi}{(n+1)}$

Cyclic: $E_k = \alpha + 2\beta \cos \frac{2k\pi}{n}$

Many predictions have been experimentally verified:

- The HOMO - LUMO gap in terms of the β constant correlates directly with the respective molecular electronic transitions observed with UV/VIS spectroscopy. For linear polyenes the energy gap is given as:

  $$\Delta E = -4\beta \sin \frac{\pi}{2(n+1)}$$

  from which a value for β can be obtained between −60 and −70 kcal/mol (−250 to −290 kJ/mol).[7]
- The predicted MO energies as stipulated by Koopmans' theorem correlate with photoelectron spectroscopy.[8]

- The Hückel delocalization energy correlates with the experimental heat of combustion. This energy is defined as the difference between the total predicted pi energy (in benzene 8β) and a hypothetical pi energy in which all ethylene units are assumed isolated each contributing 2β (making benzene 3 x 2β = 6β).
- Molecules with MO's paired up such that only the sign differs (for example α+/-β) are called **alternant hydrocarbons** and have in common small molecular dipole moments. This is in contrast to non-alternant hydrocarbons such as azulene and fulvene that have large dipole moments. The Hückel-theory is more accurate for alternant hydrocarbons.
- For cyclobutadiene the theory predicts that the two high-energy electrons occupy a degenerate pair of MO's that are neither stabilized or destabilized. Hence the square molecule would be a very reactive triplet diradical (the ground state is actually rectangular without degenerate orbitals). In fact, all cyclic conjugated hydrocarbons with a total of 4*n* pi electrons share this MO pattern and this form the basis of Hückel's rule.

Mathematics behind the Hückel method

The Hückel method can be derived from the Ritz method with a few further assumptions concerning the overlap matrix **S** and the Hamiltonian matrix **H**.

It is assumed that the overlap matrix S is the identity matrix. This means that overlap between the orbitals is neglected and the orbitals are considered orthogonal. Then the generalised eigenvalue problem of the Ritz method turns into an eigenvalue problem.

The Hamiltonian matrix **H** = (H_{ij}) is parametrised in the following way:

H_{ii} = α for C atoms and α + h_A β for other atoms A.

H_{ij} = β if the two atoms are next to each other and both C, and k_{AB} β for other neighbouring atoms A and B.

H_{ij} = 0 in any other case

The orbitals are the eigenvectors and the energies are the eigenvalues of the Hamiltonian matrix. If the substance is a pure hydrocarbon the problem can be solved without any knowledge about the parameters. For heteroatom systems, such as pyridine, values of h_A and k_{AB} have to be specified.

Hückel solution for ethylene

In the Hückel treatment for ethylene [9] , the molecular orbital Ψ is a linear combination of the 2p atomic orbitals ϕ at carbon with their ratio's c :

$$\Psi = c_1\phi_1 + c_2\phi_2$$

This equation is substituted in the Schrödinger equation:

$$H\Psi = E\Psi$$

with H the Hamiltonian and E the energy corresponding to the molecular orbital to give:

$$Hc_1\phi_1 + Hc_2\phi_2 = Ec_1\phi_1 + Ec_2\phi_2$$

This equation is multiplied by ϕ_1 and integrated to give the equation:

$$c_1(H_{11} - ES_{11}) + c_2(H_{12} - ES_{12}) = 0$$

The same equation is multiplied by ϕ_2 and integrated to give the equation:

$$c_1(H_{21} - ES_{12}) + c_2(H_{22} - ES_{22}) = 0$$

where:

$$H_{ij} = \int \phi_i H \phi_j \mathrm{d}v$$

$$S_{ij} = \int \phi_i \phi_j \mathrm{d}v$$

All diagonal Hamiltonian integrals H_{ii} are called **coulomb integrals** and those of type H_{ij}, where atoms i and j are connected, are called **resonance integrals** with these relationships:

$$H_{11} = H_{22} = \alpha$$

$$H_{12} = H_{21} = \beta$$

Other assumptions are that the overlap integral between the two atomic orbitals is 0

$$S_{11} = S_{22} = 1$$

$$S_{12} = 0$$

leading to these two homogeneous equations:

$$c_1(\alpha - E) + c_2\beta = 0$$

$$c_1\beta + c_2(\alpha - E) = 0$$

with a total of five variables. After converting this set to matrix notation:

$$\begin{bmatrix} \alpha - E & \beta \\ \beta & \alpha - E \end{bmatrix} \times \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = 0$$

the trivial solution gives both wavefunction coefficients c equal to zero which is not useful so the other (non-trivial) solution is :

$$\begin{vmatrix} \alpha - E & \beta \\ \beta & \alpha - E \end{vmatrix} = 0$$

which can be solved by expanding its determinant:

$$(\alpha - E)^2 - \beta^2 = 0$$

$$(\alpha - E)^2 = \beta^2$$

$$\alpha - E = \pm\beta$$

or

$$E = \alpha \pm \beta$$

and

$$\Psi = c_1(\phi_1 \pm \phi_2)$$

After normalization the coefficients are obtained:

$$c_1 = c_2 = \frac{1}{\sqrt{2}},$$

The constant β in the energy term is negative and therefore α + β is the lower energy corresponding to the HOMO and is α - β the LUMO energy.

External links

- Hückel method @ chem.swin.edu.au Link [10]

Further reading

- *The HMO-Model and its applications: Basis and Manipulation*, E. Heilbronner and H. Bock, English translation, 1976, Verlag Chemie.
- *The HMO-Model and its applications: Problems with Solutions*, E. Heilbronner and H. Bock, English translation, 1976, Verlag Chemie.
- *The HMO-Model and its applications: Tables of Hückel Molecular Orbitals* , E. Heilbronner and H. Bock, English translation, 1976, Verlag Chemie.

References

[1] E. Hückel, Zeitschrift für Physik, 70, 204, (1931); 72, 310, (1931); 76, 628 (1932); 83, 632, (1933)
[2] Hückel Theory for Organic Chemists, C. A. Coulson, B. O'Leary and R. B. Mallion, Academic Press,1978.
[3] *Stereochemistry of Electrocyclic Reactions* R. B. Woodward, Roald Hoffmann J. Am. Chem. Soc.; **1965**; 87(2); 395-397. doi: 10.1021/ja01080a054 (http://dx.doi.org/10.1021/ja01080a054)
[4] Andrew Streitwieser, Molecular Orbital Theory for Organic Chemists, Wiley, New York, (1961)
[5] *The chemical bond* 2nd Ed. J.N. Murrel, S.F.A. Kettle, J.M. Tedder ISBN 0471907600)
[6] *Quantum Mechanics for Organic Chemists*. Zimmerman, H., Academic Press, New York, 1975.
[7] *Use of Huckel Molecular Orbital Theory in Interpreting the Visible Spectra of Polymethine Dyes: An Undergraduate Physical Chemistry Experiment*. Bahnick, Donald A. J. Chem. Educ. **1994**, 71, 171.
[8] *Huckel theory and photoelectron spectroscopy*. von Nagy-Felsobuki, Ellak I. J. Chem. Educ. **1989**, 66, 821.
[9] *Quantum chemistry workbook* Jean-Louis Calais ISBN 0471594350
[10] http://www.chem.swin.edu.au/modules/mod3/huckel.html

Extended Huckel method

The **extended Hückel method** is a semiempirical quantum chemistry method, developed by Roald Hoffmann since 1963.[1] It is based on the Hückel method but, while the original Hückel method only considers pi orbitals, the extended method also includes the sigma orbitals.

The extended Hückel method can be used for determining the molecular orbitals, but it is not very successful in determining the structural geometry of an organic molecule. It can however determine the relative energy of different geometrical configurations. It involves calculations of the electronic interactions in a rather simple way where the electron-electron repulsions are not explicitly included and the total energy is just a sum of terms for each electron in the molecule. The off-diagonal Hamiltonian matrix elements are given by an approximation due to Wolfsberg and Helmholz that relates them to the diagonal elements and the overlap matrix element.[2]

$$H_{ij} = K S_{ij} (H_{ii} + H_{jj})/2$$

K is the Wolfsberg-Helmholtz constant, and is usually given a value of 1.75. In the extended Hückel method, only valence electrons are considered; the core electron energies and functions are supposed to be more or less constant between atoms of the same type. The method uses a series of parametrized energies calculated from atomic ionization potentials or theoretical methods to fill the diagonal of the Fock matrix. After filling the non-diagonal elements and diagonalizing the resulting Fock matrix, the energies (eigenvalues) and wavefunctions (eigenvectors) of the valence orbitals are found.

It is common in many theoretical studies to use the extended Hückel molecular orbitals as a preliminary step to determining the molecular orbitals by a more sophisticated method such as the CNDO/2 method and ab initio quantum chemistry methods. Since the EHT basis set is fixed, the monoparticle calculated wavefunctions must be projected to the basis set where the accurate calculation is to be done. One usually does this by adjusting the orbitals in the new basis to the old ones by least squares method. As only valence electron wavefunctions are found by this method, one must fill the core electron functions by orthonormalizing the rest of the basis set with the calculated orbitals and then selecting the ones with less energy. This leads to the determination of more accurate structures and electronic properties, or in the case of ab initio methods, to somewhat faster convergence.

The method was first used by Roald Hoffmann who developed, with Robert Burns Woodward, rules for elucidating reaction mechanisms (the Woodward-Hoffmann rules). He used pictures of the molecular orbitals from extended Hückel theory to work out the orbital interactions in these cycloaddition reactions.

A closely similar method was used earlier by Hoffmann and William Lipscomb for studies of boron hydrides.[3] [4] The off-diagonal Hamiltonian matrix elements were given as proportional to the overlap integral.

$$H_{ij} = K S_{ij}.$$

This simplification of the Wolfsberg and Helmholz approximation is reasonable for boron hydrides as the diagonal elements are reasonably similar due to the small difference in electronegativity between boron and hydrogen.

The method works poorly for molecules that contain atoms of very different electronegativity. To overcome this weakness, several groups have suggested iterative schemes that depend on the atomic charge. One such method, that is still widely used in inorganic and organometallic chemistry is the Fenske-Hall method. [5] [6]

A recent program for the *extended Hückel method* is YAeHMOP which stands for "yet another extended Hückel molecular orbital package".[7]

References

[1] Hoffmann, R. An Extended Hückel Theory. I. Hydrocarbons. *J. Chem. Phys* **1963**, *39*, 1397-1412. doi: 10.1063/1.1734456 (http://dx.doi.org/10.1063/1.1734456)

[2] M. Wolfsberg and L. J. Helmholz Journal of Chemical Physics, 20, 837, (1952) doi: 10.1063/1.1700580 (http://dx.doi.org/10.1063/1.1700580)

[3] R. Hoffmann and W. N. Lipscomb, Journal of Chemical Physics, 36, 2179, (1962) doi: 10.1063/1.1732849 (http://dx.doi.org/10.1063/1.1732849);37, 2872, (1962) doi: 10.1063/1.1733113 (http://dx.doi.org/10.1063/1.1733113)

[4] W. N. Lipscomb *Boron Hydrides*, W. A. Benjamin Inc., New York, 1963, Chaper 3

[5] Hall, M. B. and Fenske, R. F., Inorganic Chemistry, 11, 768 (1972)

[6] jimp2 program (http://www.chem.tamu.edu/jimp2)

[7] *Computational Chemistry*, David Young, Wiley-Interscience, 2001. Appendix A. A.3.3 pg 343, YAeHMOP

See also

- Erich Hückel
- Roald Hoffmann

Molecular graphics

Molecular graphics (**MG**) is the discipline and philosophy of studying molecules and their properties through graphical representation.[1] IUPAC limits the definition to representations on a "graphical display device".[2] Ever since Dalton's atoms and Kekule's benzene, there has been a rich history of hand-drawn atoms and molecules, and these representations have had an important influence on modern molecular graphics. This article concentrates on the use of computers to create molecular graphics. Note, however, that many molecular graphics programs and systems have close coupling between the graphics and editing commands or calculations such as in molecular modelling.

Relation to molecular models

There has been a long tradition of creating molecular models from physical materials. Perhaps the best known is Crick and Watson's model of DNA built from rods and planar sheets, but the most widely used approach is to represent all atoms and bonds explicitly using the "ball and stick" approach. This can demonstrate a wide range of properties, such as shape, relative size, and flexibility. Many chemistry courses expect that students will have access to ball and stick models. One goal of mainstream molecular graphics has been to represent the "ball and stick" model as realistically as possible and to couple this with calculations of molecular properties.

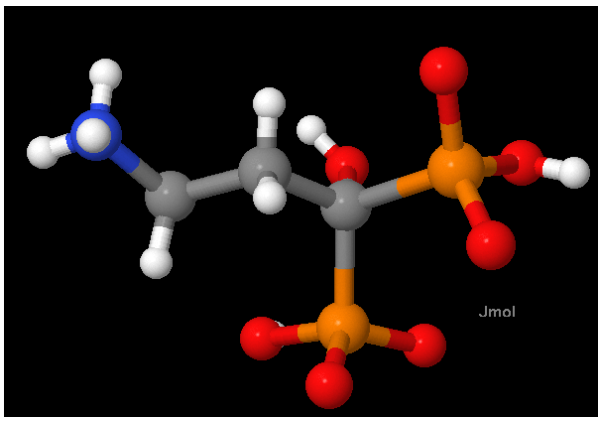


Fig. 1. Key: Hydrogen = white, carbon = grey, nitrogen = blue, oxygen = red, and phosphorus = orange.

Figure 1 shows a small molecule ($NH_3CH_2CH_2C(OH)(PO_3H)(PO_3H)$-), as drawn by the Jmol program. It is important to realise that the colours are purely a convention. Molecules can never be visible under any light microscope and atoms are not coloured, do not have hard surfaces and do not reflect light. Bonds are not rod-shaped. If physical molecular models had not existed, it is unlikely that molecular graphics would currently use this metaphor.

Comparison of physical models with molecular graphics

Physical models and computer models have partially complementary strengths and weaknesses. Physical models can be used by those without access to a computer and now can be made cheaply out of plastic materials. Their tactile and visual aspects cannot be easily reproduced by computers (although haptic devices have occasionally been built). On a computer screen, the flexibility of molecules is also difficult to appreciate; illustrating the

pseudorotation of cyclohexane is a good example of the value of mechanical models.

However, it is difficult to build large physical molecules, and all-atom physical models of even simple proteins could take weeks or months to build. Moreover, physical models are not robust and they decay over time. Molecular graphics is particularly valuable for representing global and local properties of molecules, such as electrostatic potential. Graphics can also be animated to represent molecular processes and chemical reactions, a feat that is not easy to reproduce physically.

History

Initially the rendering was on early CRT screens or through plotters drawing on paper. Molecular structures have always been an attractive choice for developing new computer graphics tools, since the input data are easy to create and the results are usually highly appealing. The first example of MG was a display of a protein molecule (Project MAC, 1966) by Cyrus Levinthal and Robert Langridge. Among the milestones in high-performance MG was the work of Nelson Max in "realistic" rendering of macromolecules using reflecting spheres.

By about 1980 many laboratories both in academia and industry had recognized the power of the computer to analyse and predict the properties of molecules, especially in materials science and the pharmaceutical industry. The discipline was often called "molecular graphics" and in 1982 a group of academics and industrialists in the UK set up the Molecular Graphics Society (MGS). Initially much of the technology concentrated either on high-performance 3D graphics, including interactive rotation or 3D rendering of atoms as spheres (sometimes with radiosity). During the 1980s a number of programs for calculating molecular properties (such as molecular dynamics and quantum mechanics) became available and the term "molecular graphics" often included these. As a result the MGS has now changed its name to the Molecular Graphics and Modelling Society (MGMS).

The requirements of macromolecular crystallography also drove MG because the traditional techniques of physical model-building could not scale. Alwyn Jones' FRODO program (and later "O") were developed to overlay the molecular electron density determined from X-ray crystallography and the hypothetical molecular structure.

Art, science and technology in molecular graphics

Both computer technology and graphic arts have contributed to molecular graphics. The development of structural biology in the 1950s led to a requirement to display molecules with thousands of atoms. The existing computer technology was limited in power, and in any case a naive depiction of all atoms left viewers overwhelmed. Most systems therefore used conventions where information was implicit or stylistic. Two vectors meeting at a point implied an atom or (in macromolecules) a complete residue (10-20 atoms).

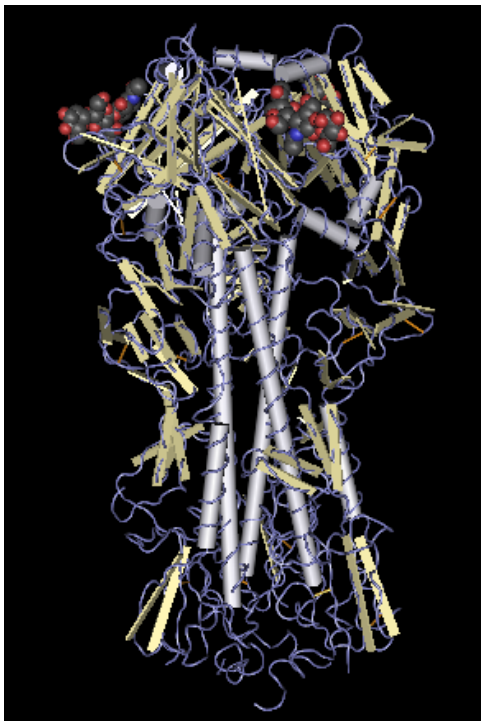


Fig. 2. Image of hemagglutinin with alpha helices depicted as cylinders and the rest of the chain as silver coils. The individual protein atoms (several thousand) have been hidden. All of the non-hydrogen atoms in the two ligands (presumably sialic acid) have been shown near the top of the diagram. Key: Carbon = grey, oxygen = red, nitrogen = blue.

The macromolecular approach was popularized by Dickerson and Geis' presentation of proteins and the graphic work of Jane Richardson through high-quality hand-drawn diagrams such as the "ribbon" representation. In this they strove to capture the intrinsic 'meaning' of the molecule. This search for the "messages in the molecule" has always accompanied the increasing power of computer graphics processing. Typically the depiction would concentrate on specific areas of the molecule (such as the active site) and this might have different colours or more detail in the number of explicit atoms or the type of depiction (e.g., spheres for atoms).

In some cases the limitations of technology have led to serendipitous methods for rendering. Most early graphics devices used vector graphics, which meant that rendering spheres and surfaces was impossible. Michael Connolly's program "MS" calculated points on the surface-accessible surface of a molecule, and the points were rendered as dots with good visibility using the new vector graphics technology, such as the Evans and Sutherland PS300 series. Thin sections ("slabs") through the structural display showed very clearly the complementarity of the surfaces for molecules binding to active sites, and the "Connolly surface" became a universal metaphor.

The relationship between the art and science of molecular graphics is shown in the exhibitions [3] sponsored by the Molecular Graphics Society. Some exhibits are created with molecular graphics programs alone, while others are collages, or involve physical materials. An example from Mike Hann (1994), inspired by Magritte's painting *Ceci n'est pas une pipe*, uses an image of a salmeterol molecule.

"*Ceci n'est pas une molecule*," writes Mike Hann, "serves to remind us that all of the graphics images presented here are not molecules, not even pictures of molecules, but pictures of icons which we believe represent some aspects of the molecule's properties."

Space-filling models

Fig. 4 is a "space-filling" representation of formic acid, where atoms are drawn to suggest the amount of space they occupy. This is necessarily an icon: in the quantum mechanical representation of molecules, there are only (positively charged) nuclei and a "cloud" of negative electrons. The electron cloud defines an approximate size for the molecule, though there can be no single precise definition of size. For many years the size of atoms has been approximated by mechanical models (CPK), where the atoms have been represented by plastic spheres whose radius (van der Waals radius) describes a sphere within which "most" of the electron density can be found. These spheres could be clicked together to show the steric aspects of the molecule rather than the positions of the nuclei. Fig. 4 shows the intricacy required to make sure that all spheres intersect correctly, and also demonstrates a reflective model.

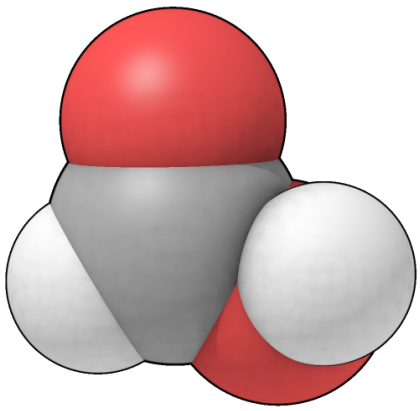


Fig. 4. Space-filling model of formic acid. Key: Hydrogen = white, carbon = black, oxygen = red.

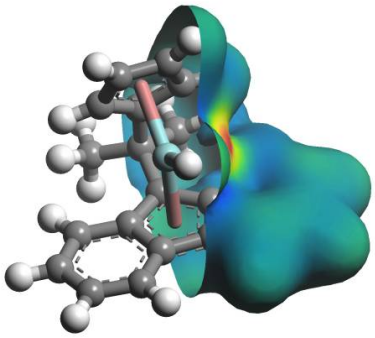


Fig. 5. A molecule (zirconocene) where part (left) is rendered as ball-and-stick and part (right) as an isosurface.

Since the atomic radii (e.g. in Fig. 4) are only slightly less than the distance between bonded atoms, the iconic spheres intersect, and in the CPK models, this was achieved by planar truncations along the bonding directions, the section being circular. When raster graphics became affordable, one of the common approaches was to replicate CPK models *in silico*. It is relatively straightforward to calculate the circles of intersection, but more complex to represent a model with hidden surface removal. A useful side product is that a conventional value for the molecular volume can be calculated.

The use of spheres is often for convenience, being limited both by graphics libraries and the additional effort required to compute complete electronic density or other space-filling quantities. It is now relatively common to see images of isosurfaces that have been coloured to show quantities such as electrostatic potential. The commonest isosurfaces are the Connolly surface, or the volume within which a given proportion of the electron density lies. The isosurface in Fig. 5 appears to show the electrostatic potential, with blue colours being negative and red/yellow (near the metal) positive. (There is no absolute convention of colouring, and red/positive, blue/negative are often confusingly reversed!) Opaque isosurfaces do not allow the atoms to be seen and identified and it is not easy to deduce them. Because of this, isosurfaces are often drawn with a degree of transparency.

Technology

Molecular graphics has always pushed the limits of display technology, and has seen a number of cycles of integration and separation of compute-host and display. Early systems like Project MAC were bespoke and unique, but in the 1970s the MMS-X and similar systems used (relatively) low-cost terminals, such as the Tektronix 4014 series, often over dial-up lines to multi-user hosts. The devices could only display static pictures but, were able to evangelize MG. In the late 1970s, it was possible for departments (such as crystallography) to afford their own hosts (e.g., PDP-11) and to attach a display (such as Evans & Sutherland's MPS) directly to the bus. The display list was kept on the host, and interactivity was good since updates were rapidly reflected in the display—at the cost of reducing most machines to a single-user system.

In the early 1980s, Evans & Sutherland (E&S) decoupled their PS300 display, which contained its own display information transformable through a dataflow architecture. Complex graphical objects could be downloaded over a serial line (e.g. 9600 baud) and then manipulated without impact on the host. The architecture was excellent for high performance display but very inconvenient for domain-specific calculations, such as electron-density fitting and energy calculations. Many crystallographers and modellers spent arduous months trying to fit such activities into this architecture.

The benefits for MG were considerable, but by the later 1980s, UNIX workstations such as Sun-3 with raster graphics (initially at a resolution of 256 by 256) had started to appear. Computer-assisted drug design in particular required raster graphics for the display of computed properties such as atomic charge and electrostatic potential. Although E&S had a high-end range of raster graphics (primarily aimed at the aerospace industry) they failed to respond to the low-end market challenge where single users, rather than engineering departments, bought workstations. As a result the market for MG displays passed to Silicon Graphics, coupled with the development of minisupercomputers (e.g., CONVEX and Alliant) which were affordable for well-supported MG laboratories. Silicon Graphics provided a graphics language, IrisGL, which was easier to use and more productive than the PS300 architecture. Commercial companies (e.g., Biosym, Polygen/MSI) ported their code to Silicon Graphics, and by the early 1990s, this was the "industry standard".

Stereoscopic displays were developed based on liquid crystal polarized spectacles, and while this had been very expensive on the PS300, it now became a commodity item. A common alternative was to add a polarizable screen to the front of the display and to provide viewers with extremely cheap spectacles with orthogonal polarization for separate eyes. With projectors such as Barco, it was possible to project stereoscopic display onto special silvered screens and supply an audience of hundreds with spectacles. In this way molecular graphics became universally known within large sectors of chemical and biochemical science, especially in the pharmaceutical industry. Because the backgrounds of many displays were black by default, it was common for modelling sessions and lectures to be held with almost all lighting turned off.

In the last decade almost all of this technology has become commoditized. IrisGL evolved to OpenGL so that molecular graphics can be run on any machine. In 1992, Roger Sayle released his RasMol program into the public domain. RasMol contained a very high-performance molecular renderer that ran on Unix/X Window, and Sayle later ported this to the Windows and Macintosh platforms. The Richardsons developed kinemages and the Mage software, which was also multi-platform. By specifying the chemical MIME type,

molecular models could be served over the Internet, so that for the first time MG could be distributed at zero cost regardless of platform. In 1995, Birkbeck College's crystallography department used this to run "Principles of Protein Structure", the first multimedia course on the Internet, which reached 100 to 200 scientists.

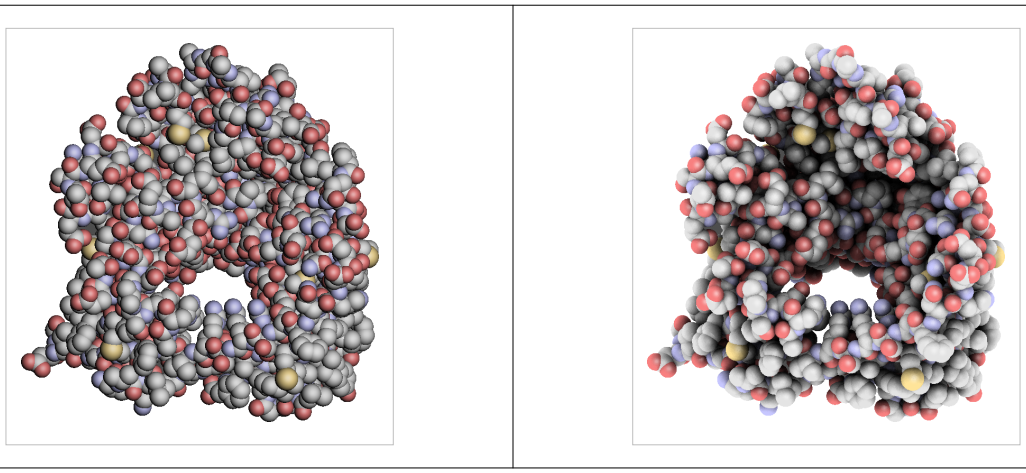


Fig. 6. A molecule of Porin (protein) shown without ambient occlusion (left) and with (right). Advanced rendering effects can improve the comprehension of the 3D shape of a molecule.

MG continues to see innovation that balances technology and art, and currently zero-cost or open source programs such as PyMOL and Jmol have very wide use and acceptance.

Recently the wide spread diffusion of advanced graphics hardware, has improved the rendering capabilities of the visualization tools. The capabilities of current shading languages allow the inclusion of advanced graphic effects (like ambient occlusion, cast shadows and non-photorealistic rendering techniques) in the interactive visualization of molecules. These graphic effects, beside being eye candy, can improve the comprehension of the three dimensional shapes of the molecules. An example of the effects that can be achieved exploiting recent graphics hardware can be seen in the simple open source visualization system QuteMol.

Algorithms

Reference frames

Drawing molecules requires a transformation between molecular coordinates (usually, but not always, in Angstrom units) and the screen. Because many molecules are chiral it is essential that the handedness of the system (almost always right-handed) is preserved. In molecular graphics the origin (0, 0) is usually at the lower left, while in many computer systems the origin is at top left. If the z-coordinate is out of the screen (towards the viewer) the molecule will be referred to right-handed axes, while the screen display will be left-handed.

Molecular transformations normally require:

- scaling of the display (but not the molecule).
- translations of the molecule and objects on the screen.
- rotations about points and lines.

Conformational changes (e.g. rotations about bonds) require rotation of one part of the molecule relative to another. The programmer must decide whether a transformation on the

screen reflects a change of view or a change in the molecule or its reference frame.

Simple

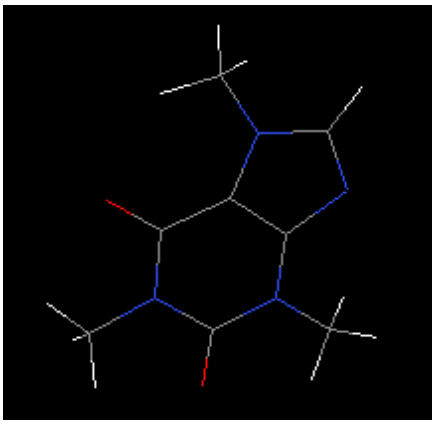


Fig. 7. Stick model of caffeine drawn in Jmol.

In early displays only vectors could be drawn e.g. (Fig. 7) which are easy to draw because no rendering or hidden surface removal is required.

On vector machines the lines would be smooth but on raster devices Bresenham's algorithm is used (note the "jaggies" on some of the bonds, which can be largely removed with antialiasing software.)

Atoms can be drawn as circles, but these should be sorted so that those with the largest z-coordinates (nearest the screen) are drawn last. Although imperfect, this often gives a reasonably attractive display. Other simple tricks which do not include hidden surface algorithms are:

- colouring each end of a bond with the same colour as the atom to which it is attached (Fig. 7).
- drawing less than the whole length of the bond (e.g. 10%-90%) to simulate the bond sticking out of a circle.
- adding a small offset white circle within the circle for an atom to simulate reflection.

Typical pseudocode for creating Fig. 7 (to fit the molecule exactly to the screen):

```
// assume:
// atoms with x, y, z coordinates (Angstrom) and elementSymbol
// bonds with pointers/references to atoms at ends
// table of colours for elementTypes
// find limits of molecule in molecule coordinates as xMin, yMin, xMax,
 yMax
scale = min(xScreenMax/(xMax-xMin), yScreenMax/(yMax-yMin))
xOffset = -xMin * scale; yOffset = -yMin * scale
for (bond in $bonds) {
  atom0 = bond.getAtom(0)
  atom1 = bond.getAtom(1)
  x0 = xOffset+atom0.getX()*scale; y0 = yOffset+atom0.getY()*scale //
(1)
  x1 = xOffset+atom1.getX()*scale; y1 = yOffset+atom1.getY()*scale //
(2)
  x1 = atom1.getX();  y1 = atom1.getY()
  xMid = (x0 + x1) /2;  yMid = (y0 + y1) /2;
  colour0 = ColourTable.getColour(atom0.getSymbol())
  drawLine (colour0, x0, y0, xMid, yMid)
  colour1 = ColourTable.getColour(atom1.getSymbol())
  drawLine (colour1, x1, y1, xMid, yMid)
}
```

Note that this assumes the origin is in the bottom left corner of the screen, with Y up the screen. Many graphics systems have the origin at the top left, with Y down the screen. In this case the lines (1) and (2) should have the y coordinate generation as:

```
y0 = yScreenMax -(yOffset+atom0.getY()*scale) // (1)
y1 = yScreenMax -(yOffset+atom1.getY()*scale) // (2)
```

Changes of this sort change the handedness of the axes so it is easy to reverse the chirality of the displayed molecule unless care is taken.

Advanced

For greater realism and better comprehension of the 3D structure of a molecule many computer graphics algorithms can be used. For many years molecular graphics has stressed the capabilities of graphics hardware and has required hardware-specific approaches. With the increasing power of machines on the desktop, portability is more important and programs such as Jmol have advanced algorithms that do not rely on hardware. On the other hand recent graphics hardware is able to interactively render very complex molecule shapes with a quality that would not be possible with standard software techniques.

Chronology

This table provides an incomplete chronology of molecular graphics advances.

Developer(s)	Approximate date	Technology	Comments
Crystallographers	< 1960	Hand-drawn	Crystal structures, with hidden atom and bond removal. Often clinographic projections.
Cyrus Levinthal, Bob Langridge	1960s	CRT	First protein display on screen (Project MAC).
Johnson, Motherwell	ca 1970	Pen plotter	ORTEP, PLUTO. Very widely deployed for publishing crystal structures.
Langridge, White, Marshall	Late 1970s	Departmental systems (PDP-11, Tektronix displays or DEC-VT11, e.g. MMS-X)	Mixture of commodity computing with early displays.
T. Alwyn Jones	1978	FRODO	Crystallographic structure solution.
Davies, Hubbard	Mid-1980s	CHEM-X, HYDRA	Laboratory systems with multicolor, raster and vector devices (Sigmex, PS300).
Biosym, Tripos, Polygen	Mid-1980s	PS300 and lower cost dumb terminals (VT200, SIGMEX)	Commercial integrated modelling and display packages.
Silicon Graphics, Sun	Late 1980s	IRIS GL (UNIX) workstations	Commodity-priced single-user workstations with stereoscopic display.
EMBL - WHAT IF [4]	1989, 2000	Machine independent	Nearly free, multifunctional, still fully supported, many free servers [5] based on it

Sayle, Richardson	1992, 1993	RasMol, Kinemage	Platform-independent MG.
MDL (van Vliet, Maffett, Adler, Holt)	1995-1998	Chime	proprietary C++ ; free browser plugin for Mac (OS9) and PCs
ChemAxon	1998-	MarvinSketch [6] & MarvinView [7]. MarvinSpace [8] (2005)	proprietary Java applet or stand-alone application.
Community efforts	2000-	Jmol, PyMol, Protein Workshop (www.pdb.org)	Open-source Java applet or stand-alone application.
San Diego Supercomputer Center	2006-	Sirius	Free for academic/non-profit institutions
NOCH	2002-	NOC [9]	Powerful and open source code molecular structure explorer
Weizmann Institute of Science - Community efforts	2008-	Proteopedia	Collaborative, 3D wiki encyclopedia of proteins & other molecules

References

[1] Dickerson, R.E.; Geis, I. (1969). *The structure and action of proteins*. Menlo Park, CA: W.A. Benjamin.
[2] International Union of Pure and Applied Chemistry (1997). " molecular graphics (http://goldbook.iupac.org/MT06970.html)". *Compendium of Chemical Terminology* Internet edition.
[3] http://www.scripps.edu/mb/goodsell/mgs_art/
[4] http://swift.cmbi.ru.nl/whatif/
[5] http://swift.cmbi.ru.nl/
[6] http://www.chemaxon.com/product/msketch.html
[7] http://www.chemaxon.com/product/mview.html
[8] http://www.chemaxon.com/product/mspace.html
[9] http://noch.sourceforge.net

See also

- Molecular Design software
- Molecular model
- Molecular modelling
- Molecular geometry
- Software for molecular mechanics modeling

External links

- The PyMOL Molecular Graphics System (http://pymol.sf.net) -- open source
  - PyMOLWiki (http://pymolwiki.org) -- community supported wiki for PyMOL
- History of Visualization of Biological Macromolecules (http://www.umass.edu/microbio/rasmol/history.htm) by Eric Martz and Eric Francoeur.
- Brief History of Molecular Mechanics/Graphics (http://stanley.chem.lsu.edu/webpub/7770-Lecture-1-intro.pdf) in LSU CHEM7770 lecture notes.
- Historical slides (http://luminary.stanford.edu/langridge/slides.htm) from Robert (Bob) Langridge. These show the influence of Crick and Watson on molecular graphics (including Levinthal's) and the development of early display technology, finishing with displays which were common in the mid-1980s on machines such as Evans and

Sutherland's PS300 series.

- Interview with Langridge. (http://luminary.stanford.edu/langridge/langridge.html) The display looking down the axis of B-DNA has been likened to a rose window.
- Nelson Max's home page (http://accad.osu.edu/~waynec/history/tree/max.html) with links to 1982 classics.
- Jmol home page (http://jmol.sourceforge.net/) contains an applet with an automatic display of many features of molecular graphics including metaphors, scripting, annotation and animation.
- Richardson Lab (http://kinemage.biochem.duke.edu/) includes Kinemage and molecular graphics images.
- History of RasMol. (http://www.openrasmol.org/history.html)
- Molecule of the Month (http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/index.html) at RCSB/PDB.
- xeo (http://sourceforge.net/projects/xeo) xeo is a free (GPL) open project management for nanostructures using Java
- Exhibitions of Molecular Graphics Art (http://www.scripps.edu/mb/goodsell/mgs_art/), 1994, 1998.
- NOCH home page (http://noch.sourceforge.net) A powerful, efficient and open source molecular graphics tool.
- eMovie (http://www.weizmann.ac.il/ISPC/eMovie.html): a tool for creation of molecular animations with PyMOL.
- Proteopedia (http://www.proteopedia.org): The collaborative, 3D encyclopedia of proteins and other molecules.
- Ascalaph Graphics (http://www.agilemolecule.com/Ascalaph/Ascalaph_Graphics.html): a molecular viewer with some geometry editing capabilities.
- Molecular Graphics and Modelling Society. (http://www.mgms.org/)
- *Journal of Molecular Graphics and Modelling* (http://www.sciencedirect.com/science?_ob=JournalURL&_cdi=5260&_auth=y&_acct=C000053194&_version=1&_urlVersion=0&_userid=1495569&md5=1e86bcce088e98890cea52f6eda84b64) (formally *Journal of Molecular Graphics*). This journal is not open access.

List of software for molecular mechanics modeling

This is a list of computer programs that are predominantly used for molecular mechanics calculations.

Min - Optimization, **MD** - Molecular Dynamics, **MC** - Monte Carlo, **QM** - Quantum mechanics. **Imp** - Implicit water. **HA** - Hardware accelerated.
Y - Yes.
I - Has interface.

Name	View 3D	Model Builder	Min	MD	MC	QM	Imp	HA	Comments	License	Website
Abalone	**Y**	**Y**	**Y**	**Y**			**Y**		Biomolecular simulations, protein folding.	Not free	Agile Molecule [1]
ACEMD[2]			**Y**	**Y**				**Y**	Molecular dynamics with CHARMM, Amber forcefields. Running on NVIDIA GPUs. Heavily optimized with CUDA.	Not free	Acellera Ltd [3]
AMBER[4]		**Y**	**Y**	**Y**			**Y**			Not free	ambermd.org [5]
Ascalaph Designer	**Y**	**Y**	**Y**	**Y**		**I**		**Y**	Molecular building (DNA, proteins, hydrocarbons, nanotubes). Molecular dynamics. GPU acceleration.	Free & Commercial	Ascalaph Project [6]
Balloon		**Y**	**Y**						2D/3D conversion and conformational analysis.	Free to use, closed source	Åbo Akademi [7]
BOSS			**Y**		**Y**	**Y**			OPLS	Commercial	Yale University [8]
CHARMM		**Y**	**Y**	**Y**	**Y**	**I**			Commercial version with multiple graphical front ends is sold by Accelrys (as CHARMm)	Not free	charmm.org [9]

ChemSketch	Y	Y	Y						Fast 2-D graphical molecule builder and 3-D viewer. Contains simplified CHARMM for fast stable inaccurate optimization of single molecules up to 1000 atoms		Advanced Chemistry Development, Inc. [10]
COSMOS	Y	Y	Y	Y	Y	I			Hybrid QM/MM COSMOS-NMR force field with fast semi-empirical calculation of electrostatic and/or NMR properties. 3-D graphical molecule builder and viewer.	Free (without GUI) and commercial	COSMOS Software [11]
Desmond				Y	Y				High Performance MD.	Free and commercial	D. E. Shaw Research [12]
GoVASP	Y		I	I		I			GoVASP is a sophisticated graphical user interface for the Vienna Ab-Initio Simulation Package (VASP). GoVASP comprises tools to prepare, perform and monitor VASP calculations and to evaluate and visualize the computed data.	Closed source/Not free/Trial available	Windiks Consulting [13]
GROMACS				Y				?	High performance MD	Free	gromacs.org [14]
GROMOS			Y	Y					Geared towards biomolecules	Not free	
LAMMPS				Y					Has potentials for soft and solid-state materials and coarse-grain systems	Free	Sandia [15]
MacroModel	Y	Y	Y	Y	Y	I	Y		OPLS-AA, GBSA solvent model, conformational sampling, minimization, MD	Not free	Schrödinger, LLC [16]

Materials Studio	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**			Materials Studio is a software environment that brings the materials simulation technology to desktop computing, solving key problems throughout the R&D process.	Closed source/Trial available	Accelrys [17]
MedeA	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**			MedeA combines leading experimental databases and major computational programs like the Vienna Ab-Initio Simulation Package (VASP) with sophisticated materials property prediction, analysis, and visualization.	Closed source/Not free	link [18]
MCCCS Towhee					**Y**				Originally designed for the prediction of fluid phase equilibria	Free	Towhee Project [19]
MDynaMix [20]				**Y**					Parallel MD	Free	Stockholm University [21]
MOE	**Y**	**Y**	**Y**	**Y**		**I**	**Y**		Molecular Operating Environment	Commercial	Chemical Computing Group [22]
MOIL	**Y**	**Y**	**Y**	**Y**					Also includes action-based algorithms (Stochastic Difference Equation in Time and Stochastic Difference Equation in Length) and locally enhanced sampling.	Free	link [23]
molecools	**Y**	**Y**							Simple Javascript molecular visualization tool		link [24]
MOLDY				**Y**					Parallel, only pair-potentials, Cell lists, modified Beeman's algorithm	Free	Moldy [25]

NAB[26]		**Y**							Generation of Models for "Unusual" DNA and RNA	Free	Case group [27]
Packmol		**Y**							Builds complex initial configurations for Molecular Dynamics		link [28]
Prime	**Y**	**Y**	**Y**		**Y**	**I**	**Y**		Homology modeling, loop and side chain optimization, minimization, OPLS-AA, SGB solvent model, parallalized		link [29]
Protein Local Optimization Program		**Y**	**Y**	**Y**	**Y**				Helix, loop, and side chain optimization. Fast energy minimization.	Not free	link [30]
QMOL	**Y**								Protein viewer	Free	DNASTAR, Inc. [31]
RasMol	**Y**								Fast viewer	Free	RasMol [32]
Raster3D	**Y**								High quality raster images	Free	University of Washington [33]
STR3DI32	**Y**	**Y**	**Y**	**Y**	''	''			Sophisticated 3-D molecule builder and viewer, advanced structural analytical algorithms, full featured molecular modeling and quantitation of stereo-electronic effects, docking and the handling of complexes.	The 200 atom version is free	Exorga, Inc. [34]
Selvita Protein Modeling Platform	**Y**	**Y**	**Y**		**Y**				Protein structure prediction, homology modeling, *ab initio* modeling, loop modeling, protein threading	Commercial	Selvita Ltd [35]
TINKER	**I**	**Y**	**Y**	**Y**	**Y**	**I**	**Y**		Software Tools for Molecular Design	Free	Washington University [36]

UCSF Chimera	**Y**	**Y**	**Y**						Visually appealing viewer, amino acid rotamers and other building, includes Antechamber and MMTK, Ambertools plugins in development.		University of California [37]
VMD + NAMD	**Y**	**Y**	**Y**	**Y**				**?**	Fast, parallel MD	Free	Beckman Institute [38]
WHAT IF	**Y**	**Y**	**I**	**I**	**I**				Visualizer for MD. Interface to GROMACS.	Not free	WHAT IF [4]
xeo	**Y**	**Y**							open project management for nanostructures		link [39]
YASARA	**Y**	**Y**		**Y**		**Y**			Molecular-graphics, -modeling and -simulation program	Not free	YASARA.org [40]
Zodiac	**Y**	**Y**	**Y**						Drug design suite		link [41]

See also

- Molecular dynamics
- Molecular Design software
- Molecule editor
- Molecular modeling on GPU
- Quantum chemistry computer programs
- List of nucleic acid simulation software
- List of protein structure prediction software
- Force field implementation

External links

- SINCRIS [42]
- Linux4Chemistry [43]
- Collaborative Computational Project [44]
- World Index of Molecular Visualization Resources [45]
- Short list of Molecular Modeling resources [46]
- OpenScience [47]
- Biological Magnetic Resonance Data Bank [48]
- Materials modelling and computer simulation codes [49]

References

[1] http://www.agilemolecule.com/Abalone/index.html
[2] M. J. Harvey, G. Giupponi and G. De Fabritiis (2009). "ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale". *Journal of Chemical THeory and Computation*: ASAP.
[3] http://www.acellera.com/index.php?arg=acemd
[4] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules". *J. Am. Chem. Soc.* **117**: 5179–5197. doi: 10.1021/ja00124a002 (http://dx.doi.org/10.1021/ja00124a002).
[5] http://ambermd.org
[6] http://www.agilemolecule.com/Ascalaph/Packages.html
[7] http://www.abo.fi/~mivainio/balloon/
[8] http://zarbi.chem.yale.edu/software.html#boss
[9] http://www.charmm.org/
[10] http://www.acdlabs.com/products/chem_dsn_lab/chemsketch/
[11] http://www.cosmos-software.de/ce_intro.html
[12] http://deshawresearch.com/resources.html
[13] http://www.govasp.com
[14] http://www.gromacs.org
[15] http://lammps.sandia.gov/
[16] http://www.schrodinger.com/ProductDescription.php?mID=6&sID=8&cID=0
[17] http://accelrys.com/products/materials-studio/
[18] http://www.materialsdesign.com/products.htm
[19] http://towhee.sourceforge.net/
[20] Lyubartsev AP, Laaksonen A (2000). "MDynaMix - A scalable portable parallel MD simulation package for arbitrary molecular mixtures". *Computer Physics Communications* **128**: 565–589. doi: 10.1016/S0010-4655(99)00529-9 (http://dx.doi.org/10.1016/S0010-4655(99)00529-9).
[21] http://www.fos.su.se/~sasha/mdynamix/
[22] http://www.chemcomp.com/
[23] http://cbsu.tc.cornell.edu/software/moil/moil.html
[24] http://blahbleh.com/molecools.php
[25] http://www.ccp5.ac.uk/moldy/moldy.html
[26] Macke T, Case DA (1998). "Modeling unusual nucleic acid structures". *Molecular Modeling of Nucleic Acids*: 379–393.
[27] http://www.scripps.edu/mb/case/
[28] http://www.ime.unicamp.br/~martinez/packmol
[29] http://www.schrodinger.com/ProductDescription.php?mID=6&sID=2&cID=0
[30] http://jacobson.compbio.ucsf.edu/plop_manual/plop_overview.htm
[31] http://www.dnastar.com/products/qmol/index.html
[32] http://www.bernstein-plus-sons.com/software/rasmol/
[33] http://skuld.bmsc.washington.edu/raster3d/raster3d.html
[34] http://www.exorga.com
[35] http://www.selvita.com/selvita-protein-modeling-platform.html
[36] http://dasher.wustl.edu/tinker/
[37] http://www.cgl.ucsf.edu/chimera/index.html
[38] http://www.ks.uiuc.edu/Research/vmd/
[39] http://sourceforge.net/projects/xeo
[40] http://www.yasara.com/
[41] http://www.zeden.org/
[42] http://ww1.iucr.org/sincris-top/logiciel/abc.html
[43] http://www.redbrick.dcu.ie/~noel/linux4chemistry/
[44] http://www.ccp14.ac.uk/mirror/mirror.htm
[45] http://molvis.sdsc.edu/visres/index.html
[46] http://www.agilemolecule.com/Software.html
[47] http://www.openscience.org/links/index.php?section=7
[48] http://www.bmrb.wisc.edu/www/software.html
[49] http://www.sklogwiki.org/SklogWiki/index.php/Category:Materials_modelling_and_computer_simulation_codes

Molecular Modeling Applications to Complex Biomolecules

Protein structure prediction

Protein structure prediction is the prediction of the three-dimensional structure of a protein from its amino acid sequence—that is, the prediction of a protein's tertiary structure from its primary structure. It is one of the most important goals pursued by bioinformatics and theoretical chemistry. Protein structure prediction is of high importance in medicine (for example, in drug design) and biotechnology (for example, in the design of novel enzymes). Every two years, the performance of current methods is assessed in the CASP experiment.

The practical role of protein structure prediction is now more important than ever. Massive amounts of protein sequence data are produced by modern large-scale DNA sequencing efforts such as the Human Genome Project. Despite community-wide efforts in structural genomics, the output of experimentally determined protein structures—typically by time-consuming and relatively expensive X-ray crystallography or NMR spectroscopy—is lagging far behind the output of protein sequences.

A number of factors exist that make protein structure prediction a very difficult task. The two main problems are that the number of possible protein structures is extremely large, and that the physical basis of protein structural stability is not fully understood. As a result, any protein structure prediction method needs a way to explore the space of possible structures efficiently (a search strategy), and a way to identify the most plausible structure (an energy function).

In comparative structure prediction (also called homology modeling), the search space is pruned by the assumption that the protein in question adopts a structure that is reasonably close to the structure of at least one known protein. In *de novo* or *ab initio* structure prediction, no such assumption is made, which results in a much harder search problem. In both cases, an energy function is needed to recognize the native structure, and to guide the search for the native structure. Unfortunately, the construction of such an energy function is to a great extent an open problem.

Direct simulation of protein folding in atomic detail, via methods such as molecular dynamics with a suitable energy function, is typically not tractable due to the high computational cost, despite the efforts of distributed computing projects such as Folding@home. Therefore, most de novo structure prediction methods rely on simplified representations of the atomic structure of proteins.

The above mentioned issues apply to all proteins, including well-behaving, small, monomeric proteins. In addition, for specific proteins (such as for example multimeric proteins and disordered proteins), the following issues also arise:

- Some proteins require stabilisation by additional domains or binding partners to adopt their native structure. This requirement is typically unknown in advance and difficult to handle by a prediction method.

- The tertiary structure of a native protein may not be readily formed without the aid of additional agents. For example, proteins known as chaperones are required for some proteins to properly fold. Other proteins cannot fold properly without modifications such as glycosylation.
- A particular protein may be able to assume multiple conformations depending on its chemical environment.
- The biologically active conformation may not be the most thermodynamically favorable.

Due to the increase in computer power, and especially new algorithms, much progress is being made to overcome these problems. However, routine de novo prediction of protein structures, even for small proteins, is still not achieved.

Ab initio protein modelling

Ab initio- or *de novo*- protein modelling methods seek to build three-dimensional protein models "from scratch", i.e., based on physical principles rather than (directly) on previously solved structures. There are many possible procedures that either attempt to mimic protein folding or apply some stochastic method to search possible solutions (i.e., global optimization of a suitable energy function). These procedures tend to require vast computational resources, and have thus only been carried out for tiny proteins. To predict protein structure *de novo* for larger proteins will require better algorithms and larger computational resources like those afforded by either powerful supercomputers (such as Blue Gene or MDGRAPE-3) or distributed computing (such as Folding@home, the Human Proteome Folding Project and Rosetta@Home). Although these computational barriers are vast, the potential benefits of structural genomics (by predicted or experimental methods) make *ab initio* structure prediction an active research field [1] .

As an intermediate step towards predicted protein structures, contact map predictions have been proposed.

Comparative protein modelling

Comparative protein modelling uses previously solved structures as starting points, or templates. This is effective because it appears that although the number of actual proteins is vast, there is a limited set of tertiary structural motifs to which most proteins belong. It has been suggested that there are only around 2000 distinct protein folds in nature, though there are many millions of different proteins.

These methods may also be split into two groups [1] :

- **Homology modeling** is based on the reasonable assumption that two homologous proteins will share very similar structures. Because a protein's fold is more evolutionarily conserved than its amino acid sequence, a target sequence can be modeled with reasonable accuracy on a very distantly related template, provided that the relationship between target and template can be discerned through sequence alignment. It has been suggested that the primary bottleneck in comparative modelling arises from difficulties in alignment rather than from errors in structure prediction given a known-good alignment.[2] Unsurprisingly, homology modelling is most accurate when the target and template have similar sequences.
- **Protein threading**[3] scans the amino acid sequence of an unknown structure against a database of solved structures. In each case, a scoring function is used to assess the

compatibility of the sequence to the structure, thus yielding possible three-dimensional models. This type of method is also known as **3D-1D fold recognition** due to its compatibility analysis between three-dimensional structures and linear protein sequences. This method has also given rise to methods performing an **inverse folding search** by evaluating the compatibility of a given structure with a large database of sequences, thus predicting which sequences have the potential to produce a given fold.

Side chain geometry prediction

Even structure prediction methods that are reasonably accurate for the peptide backbone often get the orientation and packing of the amino acid side chains wrong. Methods that specifically address the problem of predicting side chain geometry include dead-end elimination and the self-consistent mean field method. Both discretize the continuously varying dihedral angles that determine a side chain's orientation relative to the backbone into a set of rotamers with fixed dihedral angles. The methods then attempt to identify the set of rotamers that minimize the model's overall energy. Rotamers are the side chain conformations with low energy. Such methods are most useful for analyzing the protein's hydrophobic core, where side chains are more closely packed; they have more difficulty addressing the looser constraints and higher flexibility of surface residues.[4]

Protein-protein complexes

In the case of complexes of two or more proteins, where the structures of the proteins are known or can be predicted with high accuracy, protein-protein docking methods can be used to predict the structure of the complex. Information of the effect of mutations at specific sites on the affinity of the complex helps to understand the complex structure and to guide docking methods.

Software

MODELLER is a popular software tool for producing homology models using methodology derived from NMR spectroscopy data processing. SwissModel [5] provides an automated web server for basic homology modeling. I-TASSER [6] is the best server for protein structure prediction according to the recent CASP experiments [7] (CASP7 [8] and CASP8 [9]). Common software tools for protein threading are HHpred / HHsearch, bioinfo.pl [10], Robetta [11], and Phyre [12]. RAPTOR (software) is a protein threading software that is based on integer programming. The basic algorithm for threading is described in[3] and is fairly straightforward to implement. Abalone [1] is a Molecular Dynamics program for folding simulations with explicit or implicit water models.

Several distributed computing projects concerning protein structure prediction have also been implemented, such as the Folding@home, Rosetta@home, Human Proteome Folding Project, Predictor@home and TANPAKU. The Foldit program seeks to investigate the pattern-recognition and puzzle-solving abilities inherent to the human mind in order to create more successful computer protein structure prediction software.

Computational approaches provide a fast alternative route to antibody structure prediction. Recently developed antibody F_V region high resolution structure prediction algorithms like RosettaAntibody (http://antibody.graylab.jhu.edu) have been shown to generate high resolution homology models which have been used for successful docking.[13]

Reviews of software for structure prediction can be found at.[14] The progress and challenges in protein structure prediction has been reviewed in [1] .

Automatic structure prediction servers

CASP, which stands for Critical Assessment of Techniques for Protein Structure Prediction, is a community-wide experiment for protein structure prediction taking place every two years since 1994. CASP provides users and research groups with an opportunity to assess the quality of available methods and automatic servers for protein structure prediction. Official results for automatic structure prediction servers in the CASP7 benchmark (2006) are discussed by Battey *at al*.:[15] . Official CASP8 results are available here [16]. Preliminary, unofficial results for automatic servers of the recent CASP8 benchmark are summarized on several lab websites and ranked according to slightly varying criteria: Zhang lab [17], Grishin lab [18], McGuffin lab [19], Baker lab [20], Cheng lab [21]

See also

- Protein design
- Protein structure prediction software
- Protein-protein interaction prediction
- Molecular modeling software
- CASP: Annual Protein Structure Prediction Competition

References

[1] Zhang Y (2008). "Progress and challenges in protein structure prediction". *Curr Opin Struct Biol* **18** (3): 342–348. doi: 10.1016/j.sbi.2008.02.004 (http://dx.doi.org/10.1016/j.sbi.2008.02.004). Entrez Pubmed 18436442 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=18436442). PMID 18436442.

[2] Zhang Y and Skolnick J (2005). "The protein structure prediction problem could be solved using the current PDB library". *Proc Natl Acad Sci USA* **102** (4): 1029–1034. doi: 10.1073/pnas.0407152101 (http://dx.doi.org/10.1073/pnas.0407152101). Entrez Pubmed 15653774 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15653774). PMID 15653774.

[3] Bowie JU, Luthy R, Eisenberg D (1991). "A method to identify protein sequences that fold into a known three-dimensional structure". *Science* **253** (5016): 164–170. doi: 10.1126/science.1853201 (http://dx.doi.org/10.1126/science.1853201). Entrez Pubmed 1853201 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=1853201). PMID 1853201.

[4] Voigt CA, Gordon DB, Mayo SL (2000). "Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design". *J Mol Biol* **299** (3): 789–803. doi: 10.1006/jmbi.2000.3758 (http://dx.doi.org/10.1006/jmbi.2000.3758). Entrez Pubmed 10835284 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=10835284).

[5] http://swissmodel.expasy.org//SWISS-MODEL.html

[6] http://zhang.bioinformatics.ku.edu/I-TASSER

[7] http://predictioncenter.org

[8] http://www.predictioncenter.org/casp7/Casp7.html

[9] http://www.predictioncenter.org/casp8/index.cgi

[10] http://meta.bioinfo.pl/submit_wizard.pl

[11] http://robetta.bakerlab.org/

[12] http://www.sbg.bio.ic.ac.uk/phyre/index.cgi

[13] Sivasubramanian A, Sircar A, Chaudhury S, Gray J J (2009). "Toward high-resolution homology modeling of antibody Fv regions and application to antibody–antigen docking". *Proteins* **74**: 497–514. doi: 10.1002/prot.22309 (http://dx.doi.org/10.1002/prot.22309).

[14] Nayeem A, Sitkoff D, Krystek S Jr (2006). "A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models". *Protein Sci* **15**: 808–824. doi: 10.1110/ps.051892906 (http://dx.doi.org/10.1110/ps.051892906). Entrez Pubmed 16600967 (http://www.

ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16600967). PMID 16600967.
[15] Battey JN, Kopp J, Bordoli L, Read RJ, Clarke ND, Schwede T (2007). "Automated server predictions in CASP7". *Proteins* **69** (Suppl 8): 68-82.. PMID 17894354.
[16] http://predictioncenter.org/casp8/results.cgi
[17] http://zhang.bioinformatics.ku.edu/casp8/index.html
[18] http://prodata.swmed.edu/CASP8/evaluation/DomainsAll.First.html
[19] http://www.reading.ac.uk/bioinf/CASP8/index.html
[20] http://robetta.bakerlab.org/CASP8_eval_domains/
[21] http://sysbio.rnet.missouri.edu/casp8_eva/

External links

- CASP experiments home page (http://predictioncenter.org/)
- Structure Prediction Flowchart (a clickable map) (http://www.russell.embl-heidelberg.de/gtsp/flowchart2.html)

Protein design

Protein design is the design of new protein molecules from scratch, or the deliberate design of a new molecule by making calculated variations on a known structure. The number of possible amino acid sequences is enormous, but only a subset of these sequences will fold reliably and quickly to a single native state. Protein design involves identifying such sequences, in particular those with a physiologically active native state. Protein design is a rational design technique used in protein engineering.

Protein design requires an understanding of the molecular interactions that stabilize proteins in specific folded configurations fold; experience has shown, however, that protein design does not require an understanding of the dynamical process by which proteins fold. In a sense it is the reverse of structure prediction: a tertiary structure is specified, and an amino acid sequence is identified which will fold to it.

Protein design is also referred to as *inverse folding*. From a physical point of view, the native state conformation of a protein is the free energy minimum for the protein chain. Hence, designing a new protein involves the identification of the sequences which have the chosen structure as free energy minimum. This can be done by use of computer models, which, while simplifying the problem, are able to generate sequences to fold on the desired structure.

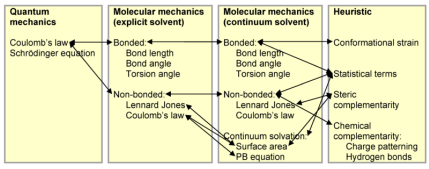
The design of minimalist computer models of proteins (lattice proteins), and the secondary structural modification of real proteins, began in the mid-1990s. The *de novo* design of real proteins became possible shortly afterwards, and the 21st century has seen the creation of small proteins with real biological function including catalysis and antiviral behaviour. There is great hope that the design of these and larger proteins will have application in medicine and bioengineering.

Computational protein design algorithms seek to identify amino acid sequences that have low energies for target structures. While the sequence-conformation space that needs to be searched is large, the most challenging requirement for computational protein design is a fast, yet accurate, energy function that can distinguish optimal sequences from similar suboptimal ones. Using computational methods, a protein with a novel fold has been designed[1], as well as sensors for un-natural molecules[2].

On the other hand, it is widely believed that not all possible protein structures are *designable*, which means that there are compact configurations of the chain which no sequences can fold to. In particular, conformations which are poor in secondary structures are unlikely to be designable. The designability of given structures is still an issue that is poorly understood.

Models of protein structure and function used in protein design

Computational protein design algorithms use models of protein energetics to evaluate how mutations would affect a protein's structure and function. These energy functions typically include a combination of molecular mechanics, knowledge-based, and other empirical terms. However, the trend has been towards using more physically based potential energy functions.[3]



Comparison of various potential energy functions

Software

EGAD: A Genetic Algorithm for protein Design[4]. A free, open-source software package for protein design and prediction of mutation effects on protein folding stabilities and binding affinities. EGAD can also consider multiple structures simultaneously for designing specific binding proteins or locking proteins into specific conformational states. In addition to natural protein residues, EGAD can also consider free-moving ligands with or without rotatable bonds. EGAD can be used with single or multiple processors.

SHARPEN [5]. A permissive open-source library for protein design and structure prediction. SHARPEN offers a variety of combinatorial optimization methods (e.g. Monte Carlo, Simulated Annealing, FASTER [6]) and can score proteins using the successful Rosetta all-atom force field or molecular mechanics force fields (OPLSaa). In addition to the protein modeling library, SHARPEN includes tools for scalable distributed computing.

WHAT IF software for protein modelling, design, validation, and visualisation.

Abalone [1] software for protein modelling and visualisation.

References

- B.I.DAHIYAT and S.L. MAYO, De Novo Protein Design: Fully Automated Sequence Selection, Science 3 October 1997:Vol. 278. no. 5335, pp. 82 – 87
- C. Sander, G. Vriend, et al., Protein Design on computers. Five new proteins: Shpilka, Grendel, Fingerclasp, Leather and Aida. *PROTEINS* 12, 105-110 (1992).
- Jin et al., *Structure*, 11, 581 (2003).
- Nagai et al., *Proc. Natl. Acad. Sci. USA*, 98, 3197 (2001).
- Saghatelian et al., *Nature*, 409, 797 (2001).
- Kuhlman et al. "Science", 302:1364 (2003)[1]
- Looger et al. "Nature", 423:185 (2003)[2]
- Pokala and Handel "J Mol Biol", 347:203 (2005)[7]

[1] http://www.sciencemag.org/cgi/content/full/302/5649/1364

[2] http://www.nature.com/nature/journal/v423/n6936/abs/nature01556.html

[3] Boas, F. E. & Harbury, P. B. (2007). "Potential energy functions for protein design." Curr. Opin. Struct. Biol. 17, 199-204. (http://www.ncbi.nlm.nih.gov/pubmed/17387014)
[4] http://egad.ucsd.edu/EGAD_manual/index.html
[5] http://www.sharp-n.com
[6] http://www.ncbi.nlm.nih.gov/pubmed/12012335
[7] http://dx.doi.org/10.1016/j.jmb.2004.12.019

See also

- PEGylation
- Protein structure prediction software
- Software for molecular modeling

Homology modeling

Homology modeling, also known as **comparative modeling** of protein refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.[1]

The quality of the homology model is dependent on the quality of the sequence alignment and template structure. The approach can be complicated by the presence of alignment gaps (commonly called indels) that indicate a structural region present in the target but not in the template, and by structure gaps in the template that arise from poor resolution in the experimental procedure (usually X-ray crystallography) used to solve the structure. Model quality declines with decreasing sequence identity; a typical model has ~1-2 Å root mean square deviation between the matched C^{α} atoms at 70% sequence identity but only 2-4 Å agreement at 25% sequence identity. However, the errors are significantly higher in the loop regions, where the amino acid sequences of the target and template proteins may be completely different.

Regions of the model that were constructed without a template, usually by loop modeling, are generally much less accurate than the rest of the model. Errors in side chain packing and position also increase with decreasing identity, and variations in these packing configurations have been suggested as a major reason for poor model quality at low identity.[2] Taken together, these various atomic-position errors are significant and impede the use of homology models for purposes that require atomic-resolution data, such as drug design and protein-protein interaction predictions; even the quaternary structure of a protein may be difficult to predict from homology models of its subunit(s). Nevertheless, homology models can be useful in reaching *qualitative* conclusions about the biochemistry of the query sequence, especially in formulating hypotheses about why certain residues are conserved, which may in turn lead to experiments to test those hypotheses. For example, the spatial arrangement of conserved residues may suggest whether a particular residue is conserved to stabilize the folding, to participate in binding some small molecule, or to

foster association with another protein or nucleic acid.

Homology modeling can produce high-quality structural models when the target and template are closely related, which has inspired the formation of a structural genomics consortium dedicated to the production of representative experimental structures for all classes of protein folds.[3] The chief inaccuracies in homology modeling, which worsen with lower sequence identity, derive from errors in the initial sequence alignment and from improper template selection.[4] Like other methods of structure prediction, current practice in homology modeling is assessed in a biannual large-scale experiment known as the Critical Assessment of Techniques for Protein Structure Prediction, or CASP.

Motivation

The method of homology modeling is based on the observation that protein tertiary structure is better conserved than amino acid sequence.[1] Thus, even proteins that have diverged appreciably in sequence but still share detectable similarity will also share common structural properties, particularly the overall fold. Because it is difficult and time-consuming to obtain experimental structures from methods such as X-ray crystallography and protein NMR for every protein of interest, homology modeling can provide useful structural models for generating hypotheses about a protein's function and directing further experimental work.

There are exceptions to the general rule that proteins sharing significant sequence identity will share a fold. For example, a judiciously chosen set of mutations of less than 50% of a protein can cause the protein to adopt a completely different fold.[5] [6] However, such a massive structural rearrangement is unlikely to occur in evolution, especially since the protein is usually under the constraint that it must fold properly and carry out its function in the cell. Consequently, the roughly folded structure of a protein (its "topology") is conserved longer than its amino-acid sequence and much longer than the corresponding DNA sequence; in other words, two proteins may share a similar fold even if their evolutionary relationship is so distant that it cannot be discerned reliably. For comparison, the function of a protein is conserved much *less* than the protein sequence, since relatively few changes in amino-acid sequence are required to take on a related function.

Steps in model production

The homology modeling procedure can be broken down into four sequential steps: template selection, target-template alignment, model construction, and model assessment.[1] The first two steps are often essentially performed together, as the most common methods of identifying templates rely on the production of sequence alignments; however, these alignments may not be of sufficient quality because database search techniques prioritize speed over alignment quality. These processes can be performed iteratively to improve the quality of the final model, although quality assessments that are not dependent on the true target structure are still under development.

Optimizing the speed and accuracy of these steps for use in large-scale automated structure prediction is a key component of structural genomics initiatives, partly because the resulting volume of data will be too large to process manually and partly because the goal of structural genomics requires providing models of reasonable quality to researchers who are not themselves structure prediction experts.[1]

Template selection and sequence alignment

The critical first step in homology modeling is the identification of the best template structure, if indeed any are available. The simplest method of template identification relies on serial pairwise sequence alignments aided by database search techniques such as FASTA and BLAST. More sensitive methods based on multiple sequence alignment - of which PSI-BLAST is the most common example - iteratively update their position-specific scoring matrix to successively identify more distantly related homologs. This family of methods has been shown to produce a larger number of potential templates and to identify better templates for sequences that have only distant relationships to any solved structure. Protein threading, also known as fold recognition or 3D-1D alignment, can also be used as a search technique for identifying templates to be used in traditional homology modeling methods.[1] When performing a BLAST search, a reliable first approach is to identify hits with a sufficiently low *E*-value, which are considered sufficiently close in evolution to make a reliable homology model. Other factors may tip the balance in marginal cases; for example, the template may have a function similar to that of the query sequence, or it may belong to a homologous operon. However, a template with a poor *E*-value should generally not be chosen, even if it is the only one available, since it may well have a wrong structure, leading to the production of a misguided model. A better approach is to submit the primary sequence to fold-recognition servers or, better still, consensus meta-servers which improve upon individual fold-recognition servers by identifying similarities (consensus) among independent predictions.

Often several candidate template structures are identified by these approaches. Although some methods can generate hybrid models from multiple templates, most methods rely on a single template. Therefore, choosing the best template from among the candidates is a key step, and can affect the final accuracy of the structure significantly. This choice is guided by several factors, such as the similarity of the query and template sequences, of their functions, and of the predicted query and observed template secondary structures. Perhaps most importantly, the *coverage* of the aligned regions: the fraction of the query sequence structure that can be predicted from the template, and the plausibility of the resulting model. Thus, sometimes several homology models are produced for a single query sequence, with the most likely candidate chosen only in the final step.

It is possible to use the sequence alignment generated by the database search technique as the basis for the subsequent model production; however, more sophisticated approaches have also been explored. One proposal generates an ensemble of stochastically defined pairwise alignments between the target sequence and a single identified template as a means of exploring "alignment space" in regions of sequence with low local similarity.[7] "Profile-profile" alignments that first generate a sequence profile of the target and systematically compare it to the sequence profiles of solved structures; the coarse-graining inherent in the profile construction is thought to reduce noise introduced by sequence drift in nonessential regions of the sequence.[8]

Model generation

Given a template and an alignment, the information contained therein must be used to generate a three-dimensional structural model of the target, represented as a set of Cartesian coordinates for each atom in the protein. Three major classes of model generation methods have been proposed.[9]

Fragment assembly

The original method of homology modeling relied on the assembly of a complete model from conserved structural fragments identified in closely related solved structures. For example, a modeling study of serine proteases in mammals identified a sharp distinction between "core" structural regions conserved in all experimental structures in the class, and variable regions typically located in the loops where the majority of the sequence differences were localized. Thus unsolved proteins could be modeled by first constructing the conserved core and then substituting variable regions from other proteins in the set of solved structures.[10] Current implementations of this method differ mainly in the way they deal with regions that are not conserved or that lack a template.[11]

Segment matching

The segment-matching method divides the target into a series of short segments, each of which is matched to its own template fitted from the Protein Data Bank. Thus, sequence alignment is done over segments rather than over the entire protein. Selection of the template for each segment is based on sequence similarity, comparisons of alpha carbon coordinates, and predicted steric conflicts arising from the van der Waals radii of the divergent atoms between target and template. [12]

Satisfaction of spatial restraints

The most common current homology modeling method takes its inspiration from calculations required to construct a three-dimensional structure from data generated by NMR spectroscopy. One or more target-template alignments are used to construct a set of geometrical criteria that are then converted to probability density functions for each restraint. Restraints applied to the main protein internal coordinates - protein backbone distances and dihedral angles - serve as the basis for a global optimization procedure that originally used conjugate gradient energy minimization to iteratively refine the positions of all heavy atoms in the protein.[13]

This method had been dramatically expanded to apply specifically to loop modeling, which can be extremely difficult due to the high flexibility of loops in proteins in aqueous solution.[14] A more recent expansion applies the spatial-restraint model to electron density maps derived from cryoelectron microscopy studies, which provide low-resolution information that is not usually itself sufficient to generate atomic-resolution structural models.[15] To address the problem of inaccuracies in initial target-template sequence alignment, an iterative procedure has also been introduced to refine the alignment on the basis of the initial structural fit.[16] The most commonly used software in spatial restraint-based modeling is MODELLER and a database called ModBase has been established for reliable models generated with it.[17]

Loop modeling

Regions of the target sequence that are not aligned to a template are modeled by loop modeling; they are the most susceptible to major modeling errors and occur with higher frequency when the target and template have low sequence identity. The coordinates of unmatched sections determined by loop modeling programs are generally much less accurate than those obtained from simply copying the coordinates of a known structure, particularly if the loop is longer than 10 residues. The first two sidechain dihedral angles (χ_1 and χ_2) can usually be estimated within 30° for an accurate backbone structure; however, the later dihedral angles found in longer side chains such as lysine and arginine are notoriously difficult to predict. Moreover, small errors in χ_1 (and, to a lesser extent, in χ_2) can cause relatively large errors in the positions of the atoms at the terminus of side chain; such atoms often have a functional importance, particularly when located near the active site.

Model assessment

Assessment of homology models without reference to the true target structure is usually performed with two methods: statistical potentials or physics-based energy calculations. Both methods produce an estimate of the energy (or an energy-like analog) for the model or models being assessed; independent criteria are needed to determine acceptable cutoffs. Neither of the two methods correlates exceptionally well with true structural accuracy, especially on protein types underrepresented in the PDB, such as membrane proteins.

Statistical potentials are empirical methods based on observed residue-residue contact frequencies among proteins of known structure in the PDB. They assign a probability or energy score to each possible pairwise interaction between amino acids and combine these pairwise interaction scores into a single score for the entire model. Some such methods can also produce a residue-by-residue assessment that identifies poorly scoring regions within the model, though the model may have a reasonable score overall.[18] These methods emphasize the hydrophobic core and solvent-exposed polar amino acids often present in globular proteins. Examples of popular statistical potentials include Prosa and DOPE. Statistical potentials are more computationally efficient than energy calculations.[18]

Physics-based energy calculations aim to capture the interatomic interactions that are physically responsible for protein stability in solution, especially van der Waals and electrostatic interactions. These calculations are performed using a molecular mechanics force field; proteins are normally too large even for semi-empirical quantum mechanics-based calculations. The use of these methods is based on the energy landscape hypothesis of protein folding, which predicts that a protein's native state is also its energy minimum. Such methods usually employ implicit solvation, which provides a continuous approximation of a solvent bath for a single protein molecule without necessitating the explicit representation of individual solvent molecules. A force field specifically constructed for model assessment is known as the Effective Force Field (EFF) and is based on atomic parameters from CHARMM.[19]

A very extensive model validation report can be obtained using the Radboud Universiteit Nijmegen [20] "*What Check*" software which is one option of the Radboud Universiteit Nijmegen [4] "*What If*" software package; it produces a many page document with extensive analyses of nearly 200 scientific and administrative aspects of the model. "*What Check*" is available as a free server [5]; it can also be used to validate experimentally determined

structures of macromolecules.

One newer method for model assessment relies on machine learning techniques such as neural nets, which may be trained to assess the structure directly or to form a consensus among multiple statistical and energy-based methods. Very recent results using support vector machine regression on a jury of more traditional assessment methods outperformed common statistical, energy-based, and machine learning methods.[21]

Structural comparison methods

The assessment of homology models' accuracy is straightforward when the experimental structure is known. The most common method of comparing two protein structures uses the root-mean-square deviation (RMSD) metric to measure the mean distance between the corresponding atoms in the two structures after they have been superimposed. However, RMSD does underestimate the accuracy of models in which the core is essentially correctly modeled, but some flexible loop regions are inaccurate.[22] A method introduced for the modeling assessment experiment CASP is known as the global distance test (GDT) and measures the total number of atoms whose distance from the model to the experimental structure lies under a certain distance cutoff.[22] Both methods can be used for any subset of atoms in the structure, but are often applied to only the alpha carbon or protein backbone atoms to minimize the noise created by poorly modeled side chain rotameric states, which most modeling methods are not optimized to predict.[23]

Benchmarking

Several large-scale benchmarking efforts have been made to assess the relative quality of various current homology modeling methods. CASP is a community-wide prediction experiment that runs every two years during the summer months and challenges prediction teams to submit structural models for a number of sequences whose structures have recently been solved experimentally but have not yet been published. Its partner CAFASP has run in parallel with CASP but evaluates only models produced via fully automated servers. Continuously running experiments that do not have prediction 'seasons' focus mainly on benchmarking publicly available webservers. LiveBench and EVA run continuously to assess participating servers' performance in prediction of imminently released structures from the PDB. CASP and CAFASP serve mainly as evaluations of the state of the art in modeling, while the continuous assessments seek to evaluate the model quality that would be obtained by a non-expert user employing publicly available tools.

Accuracy

The accuracy of the structures generated by homology modeling is highly dependent on the sequence identity between target and template. Above 50% sequence identity, models tend to be reliable, with only minor errors in side chain packing and rotameric state, and an overall RMSD between the modeled and the experimental structure falling around 1 Â. This error is comparable to the typical resolution of a structure solved by NMR. In the 30-50% identity range, errors can be more severe and are often located in loops. Below 30% identity, serious errors occur, sometimes resulting in the basic fold being mis-predicted.[9] This low-identity region is often referred to as the "twilight zone" within which homology modeling is extremely difficult, and to which it is possibly less suited than fold recognition methods.[24]

At high sequence identities, the primary source of error in homology modeling derives from the choice of the template or templates on which the model is based, while lower identities exhibit serious errors in sequence alignment that inhibit the production of high-quality models.[4] It has been suggested that the major impediment to quality model production is inadequacies in sequence alignment, since "optimal" structural alignments between two proteins of known structure can be used as input to current modeling methods to produce quite accurate reproductions of the original experimental structure.[25]

Attempts have been made to improve the accuracy of homology models built with existing methods by subjecting them to molecular dynamics simulation in an effort to improve their RMSD to the experimental structure. However, current force field parameterizations may not be sufficiently accurate for this task, since homology models used as starting structures for molecular dynamics tend to produce slightly worse structures.[26] Slight improvements have been observed in cases where significant restraints were used during the simulation.[27]

Sources of error

The two most common and large-scale sources of error in homology modeling are poor template selection and inaccuracies in target-template sequence alignment.[4] [28] Controlling for these two factors by using a structural alignment, or a sequence alignment produced on the basis of comparing two solved structures, dramatically reduces the errors in final models; these "gold standard" alignments can be used as input to current modeling methods to produce quite accurate reproductions of the original experimental structure.[25] Results from the most recent CASP experiment suggest that "consensus" methods collecting the results of multiple fold recognition and multiple alignment searches increase the likelihood of identifying the correct template; similarly, the use of multiple templates in the model-building step may be less optimal than the use of the single correct template but more optimal than the use of a single suboptimal one.[28] Alignment errors may be minimized by the use of a multiple alignment even if only one template is used, and by the iterative refinement of local regions of low similarity.[1] [7] A lesser source of model errors are errors in the template structure. The http://swift.cmbi.ru.nl/gv/pdbreport/ PDBREPORT database lists several million, mostly very small but occasionally dramatic, errors in experimental (template) structures that have been deposited in the PDB.

Serious local errors can arise in homology models where an insertion or deletion mutation or a gap in a solved structure result in a region of target sequence for which there is no corresponding template. This problem can be minimized by the use of multiple templates, but the method is complicated by the templates' differing local structures around the gap and by the likelihood that a missing region in one experimental structure is also missing in other structures of the same protein family. Missing regions are most common in loops where high local flexibility increases the difficulty of resolving the region by structure-determination methods. Although some guidance is provided even with a single template by the positioning of the ends of the missing region, the longer the gap, the more difficult it is to model. Loops of up to about 9 residues can be modeled with moderate accuracy in some cases if the local alignment is correct.[1] Larger regions are often modeled individually using ab initio structure prediction techniques, although this approach has met with only isolated success.[29]

The rotameric states of side chains and their internal packing arrangement also present difficulties in homology modeling, even in targets for which the backbone structure is relatively easy to predict. This is partly due to the fact that many side chains in crystal structures are not in their "optimal" rotameric state as a result of energetic factors in the hydrophobic core and in the packing of the individual molecules in a protein crystal.[30] One method of addressing this problem requires searching a rotameric library to identify locally low-energy combinations of packing states.[31] It has been suggested that a major reason that homology modeling so difficult when target-template sequence identity lies below 30% is that such proteins have broadly similar folds but widely divergent side chain packing arrangements.[2]

Utility

Uses of the structural models include protein-protein interaction prediction, protein-protein docking, molecular docking, and functional annotation of genes identified in an organism's genome.[32] Even low-accuracy homology models can be useful for these purposes, because their inaccuracies tend to be located in the loops on the protein surface, which are normally more variable even between closely related proteins. The functional regions of the protein, especially its active site, tend to be more highly conserved and thus more accurately modeled.[9]

Homology models can also be used to identify subtle differences between related proteins that have not all been solved structurally. For example, the method was used to identify cation binding sites on the Na^+/K^+ ATPase and to propose hypotheses about different ATPases' binding affinity.[33] Used in conjunction with molecular dynamics simulations, homology models can also generate hypotheses about the kinetics and dynamics of a protein, as in studies of the ion selectivity of a potassium channel.[34] Large-scale automated modeling of all identified protein-coding regions in a genome has been attempted for the yeast *Saccharomyces cerevisiae*, resulting in nearly 1000 quality models for proteins whose structures had not yet been determined at the time of the study, and identifying novel relationships between 236 yeast proteins and other previously solved structures.[35]

See also

- Protein structure prediction
- Protein structure prediction software
- Protein threading

References

[1] Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.

[2] Chung SY, Subbiah S. (1996.) A structural explanation for the twilight zone of protein sequence homology. *Structure* 4: 1123–27.

[3] Williamson AR. (2000). Creating a structural genomics consortium. *Nat Struct Biol* 7 S1(11s):953.

[4] Venclovas C, Margeleviĉius M. (2005). Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. *Proteins* 61(S7):99-105.

[5] Dalal S, Balasubramanian S, Regan L. (1997). Transmuting alpha helices and beta sheets. *Fold Des* 2(5):R71-9.

[6] Dalal S, Balasubramanian S, Regan L. (1997). Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol* 4(7):548-52.

[7] Muckstein U, Hofacker IL, Stadler PF. (2002). Stochastic pairwise alignments. *Bioinformatics* 18 Suppl 2:S153-60.
[8] Rychlewski L, Zhang B, Godzik A. (1998). Fold and function predictions for Mycoplasma genitalium proteins. *Fold Des* 3(4):229-38.
[9] Baker D, Sali A. (2001). Protein structure prediction and structural genomics. *Science* 294(5540):93-96.
[10] Greer J. (1981). Comparative model-building of the mammalian serine proteases 153(4):1027-42.
[11] Wallner B, Elofsson A. (2005). All are not equal: A benchmark of different homology modeling programs. *Protein Science* 14:1315-1327.
[12] Levitt M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol* 226(2): 507-33.
[13] Sali A, Blundell TL. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779-815.
[14] Fiser A, Sali A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 19(18):2500-1.
[15] Topf M, Baker ML, Marti-Renom MA, Chiu W, Sali A. (2006). Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J Mol Biol* 357(5):1655-68.
[16] John B, Sali A. (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res* 31(14):3982-92.
[17] Ursula Pieper, Narayanan Eswar, Hannes Braberg, M.S. Madhusudhan, Fred Davis, Ashley C. Stuart, Nebojsa Mirkovic, Andrea Rossi, Marc A. Marti-Renom, Andras Fiser, Ben Webb, Daniel Greenblatt, Conrad Huang, Tom Ferrin, Andrej Sali. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res* 32, D217-D222, 2004.
[18] Sippl MJ. (1993). Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins* 17:355-62.
[19] Lazaridis T. and Karplus M. 1999a. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* 288: 477–487
[20] http://swift.cmbi.ru.nl/gv/whatcheck/
[21] Eramian D, Shen M, Devos D, Melo F, Sali A, Marti-Renom MA. (2006). A composite score for predicting errors in protein structure models. *Protein Science* 15:1653-1666.
[22] Zemla A. (2003). LGA - A Method for Finding 3-D Similarities in Protein Structures. Nucleic Acids Research, 31(13):3370-3374.
[23] Mount DM. (2004). *Bioinformatics: Sequence and Genome Analysis* 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
[24] Blake JD, Cohen FE. (2001). Pairwise sequence alignment below the twilight zone. *J Mol Biol* 307(2):721-35.
[25] Zhang Y and Skolnick J. (2005). The protein structure prediction problem could be solved using the current PDB library. Proc. Natl. Acad. Sci. USA 102(4):1029-34.
[26] Koehl P, Levitt M. (1999). A brighter future for protein structure prediction. *Nat Struct Biol* 6(2):108-11.
[27] Flohil JA, Vriend G, Berendsen HJ. (2002). Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. *Proteins* 48(4):593-604.
[28] Ginalski K. (2006). Comparative modeling for protein structure prediction. *Curr Opin Struct Biol* 16(2):172-7.
[29] Kryshtafovych A, Venclovas C, Fidelis K, Moult J. (2005). Progress over the first decade of CASP experiments. *Proteins* 61(S7):225-36.
[30] Vasquez M. (1996). Modeling side-chain conformation. *Curr Opin Struct Biol* 6(2):217-21.
[31] Wilson C, Gregoret LM, Agard DA. (1993). Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. *J Mol Biol* 229(4):996-1006.
[32] Gopal S, Schroeder M, Pieper U, Sczyrba A, Aytekin-Kurban G, Bekiranov S, Fajardo JE, Eswar N, Sanchez R, Sali A, Gaasterland T. (2001). Homology-based annotation yields 1,042 new candidate genes in the Drosophila melanogaster genome. *Nat Genet* 27(3):337-40.
[33] Ogawa H, Toyoshima C. (2002). Homology modeling of the cation binding sites of Na+K+-ATPase. *Proc Natl Acad Sci USA* 99(25):15977-15982
[34] Capener CE, Shrivastava IH, Ranatunga KM, Forrest LR, Smith GR, Sansom MSP. (2000). Homology Modeling and Molecular Dynamics Simulation Studies of an Inward Rectifier Potassium Channel. *Biophys J* 78(6):2929-2942
[35] Sánchez R, Sali A. (1998). Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. *Proc Natl Acad Sci USA* 95(23):13597-13602.

Loop modeling

Loop modeling is a problem in protein structure prediction requiring the prediction of the conformations of loop regions in proteins without the use of a structural template. The problem arises often in homology modeling, where the tertiary structure of an amino acid sequence is predicted based on a sequence alignment to a *template*, or a second sequence whose structure is known. Because loops have highly variable sequences even within a given structural motif or protein fold, they often correspond to unaligned regions in sequence alignments; they also tend to be located at the solvent-exposed surface of globular proteins and thus are more conformationally flexible. Consequently, they often cannot be modeled using standard homology modeling techniques. More constrained versions of loop modeling are also used in the data fitting stages of solving a protein structure by X-ray crystallography, because loops can correspond to regions of low electron density and are therefore difficult to resolve.

Regions of a structural model that were predicted by loop modeling tend to be much less accurate than regions that were predicted using template-based techniques. The extent of the inaccuracy increases with the number of amino acids in the loop. The loop amino acids' side chains dihedral angles are often approximated from a rotamer library, but can worsen the inaccuracy of side chain packing in the overall model. Andrej Sali's homology modeling suite MODELLER includes a facility explicitly designed for loop modeling by a satisfaction of spatial restraints method.

Short loops

In general, the most accurate predictions are for loops of fewer than 8 amino acids. Extremely short loops of three residues can be determined from geometry alone, provided that the bond lengths and bond angles are specified. Slightly longer loops are often determined from a "spare parts" approach, in which loops of similar length are taken from known crystal structures and adapted to the geometry of the flanking segments. In some methods, the bond lengths and angles of the loop region are allowed to vary, in order to obtain a better fit; in other cases, the constraints of the flanking segments may be varied to find more "protein-like" loop conformations. The accuracy of such short loops may be almost as accurate as that of the homology model upon which it is based. It should also be considered that the loops in proteins may not be well-structured and therefore have no one conformation that could be predicted; NMR experiments indicate that solvent-exposed loops are "floppy" and adopt many conformations, while the loop conformations seen by X-ray crystallography may merely reflect crystal packing interactions, or the stabilizing influence of crystallization co-solvents.

References

- Mount DM. (2004). Bioinformatics: Sequence and Genome Analysis 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
- Chung SY, Subbiah S. (1996.) A structural explanation for the twilight zone of protein sequence homology. Structure 4: 1123–27.

External links

- MODLOOP [1], public server for access to MODELLER's loop modeling facility

References

[1] http://modbase.compbio.ucsf.edu/modloop

MODELLER

MODELLER is a computer program used in producing homology models of protein tertiary structures as well as quaternary structures (rarer). It implements a technique inspired by nuclear magnetic resonance known as *satisfaction of spatial restraints*, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The method relies on an input sequence alignment between the target amino acid sequence to be modeled and a template protein whose structure has been solved.

The program also incorporates limited functionality for ab initio structure prediction of loop regions of proteins, which are often highly variable even among homologous proteins and therefore difficult to predict by homology modeling.

MODELLER was originally written and is currently maintained by Andrej Sali at the University of California, San Francisco. Although it is freely available for academic use, graphical user interfaces and commercial versions are distributed by Accelrys.

External links

- MODELLER [1]

References

- Sali A, Blundell TL. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.
- Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291-325.
- Fiser A, Sali A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374:461-91

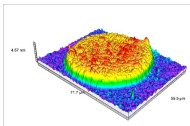
References

[1] http://salilab.org/modeller/

Molecular models of DNA

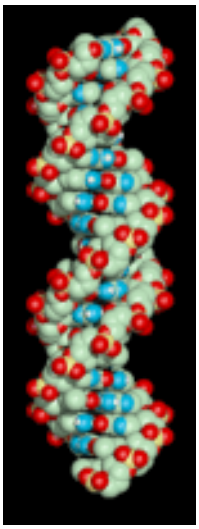
Molecular models of DNA structures are representations of the molecular geometry and topology of Deoxyribonucleic acid (DNA) molecules using one of several means, such as: closely packed spheres (CPK models) made of plastic, metal wires for 'skeletal models', graphic computations and animations by computers, artistic rendering, and so on, with the aim of simplifying and presenting the essential, physical and chemical, properties of DNA molecular structures either *in vivo* or *in vitro*. Computer molecular models also allow animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*. Thus, an old standing dynamic problem is how DNA "self-replication" takes place in living cells that should involve transient uncoiling of supercoiled DNA fibers. Although DNA consists of relatively rigid, very large elongated biopolymer molecules called "fibers" or chains (that are made of repeating nucleotide units of four basic types, attached to deoxyribose and phosphate groups), its molecular structure *in vivo* undergoes dynamic configuration changes that involve dynamically attached water molecules and ions. Supercoiling, packing with histones in chromosome structures, and other such supramolecular aspects also involve *in vivo* DNA topology which is even more complex than DNA molecular geometry, thus turning molecular modeling of DNA into an especially challenging problem for both molecular biologists and biotechnologists. Like other large molecules and biopolymers, DNA often exists in multiple stable geometries (that is, it exhibits conformational isomerism) and configurational, quantum states which are close to each other in energy on the potential energy surface of the DNA molecule. Such geometries can also be computed, at least in principle, by employing *ab initio* quantum chemistry methods that have high accuracy for small molecules. Such quantum geometries define an important class of *ab initio* molecular models of DNA whose exploration has barely started.

In an interesting twist of roles, the DNA molecule itself was proposed to be utilized for quantum computing. Both DNA nanostructures as well as DNA 'computing' biochips have been built (see biochip image at right).



DNA computing biochip:3D

The more advanced, computer-based molecular models of DNA involve molecular dynamics simulations as well as quantum mechanical computations of vibro-rotations, delocalized molecular orbitals (MOs), electric dipole moments, hydrogen-bonding, and so on.



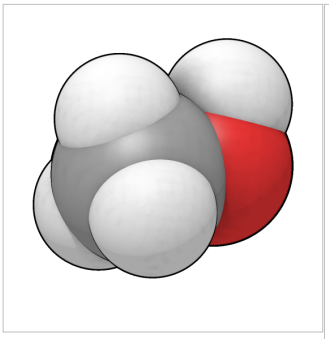
Spinning DNA generic model.

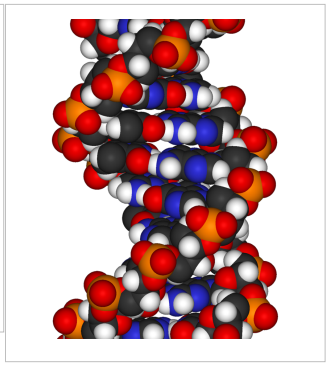
Importance

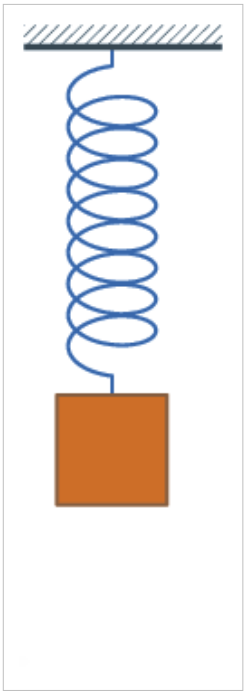
From the very early stages of structural studies of DNA by X-ray diffraction and biochemical means, molecular models such as the Watson-Crick double-helix model were successfully employed to solve the 'puzzle' of DNA structure, and also find how the latter relates to its key functions in living cells. The first high quality X-ray diffraction patterns of A-DNA were reported by Rosalind Franklin and Raymond Gosling in 1953[1] . The first calculations of the Fourier transform of an atomic helix were reported one year earlier by Cochran, Crick and Vand [2] , and were followed in 1953 by the computation of the Fourier transform of a coiled-coil by Crick[3] . The first reports of a double-helix molecular model of B-DNA structure were made by Watson and Crick in 1953[4] [5] . Last-but-not-least, Maurice F. Wilkins, A. Stokes and H.R. Wilson, reported the first X-ray patterns of *in vivo* B-DNA in partially oriented salmon sperm heads [6] . The development of the first correct double-helix molecular model of DNA by Crick and Watson may not have been possible without the biochemical evidence for the nucleotide base-pairing ([A---T]; [C---G]), or Chargaff's rules[7] [8] [9] [10] [11] [12] .

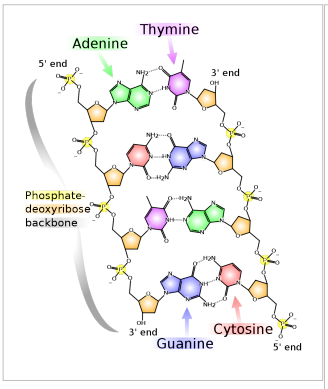
Examples of DNA molecular models

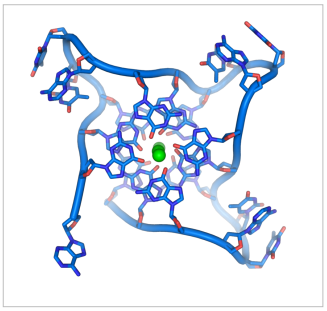
Animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. The first DNA model is a space-filling, or CPK, model of the DNA double-helix whereas the third is an animated wire, or skeletal type, molecular model of DNA. The last two DNA molecular models in this series depict quadruplex DNA [13] that may be involved in certain cancers[14] [15] . The last figure on this panel is a molecular model of hydrogen bonds between water molecules in ice that are similar to those found in DNA.

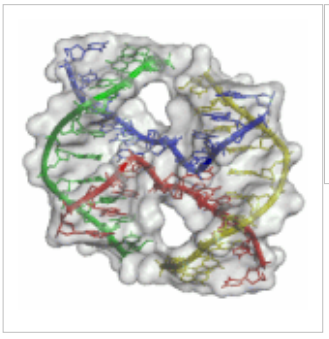


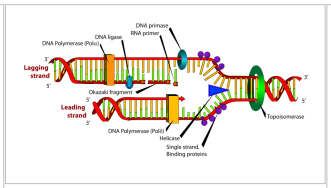


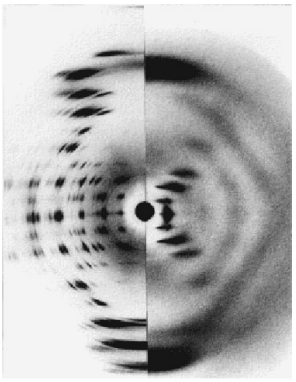



Thymine
Adenine
5' end
3' end
Phosphate-
deoxyribose
backbone
3' end
Guanine
Cytosine
5' end

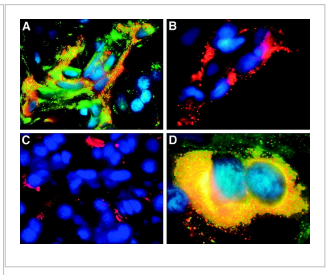


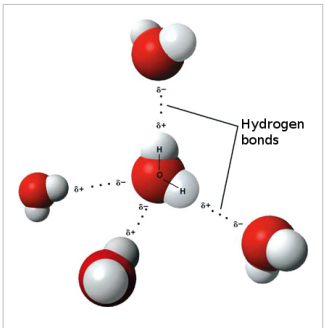



Lagging strand
Leading strand
DNA Polymerase (Pol I)
DNA ligase
DNA primase
RNA primer
Okazaki fragment
DNA Polymerase (Pol III)
Helicase
Single strand. Binding proteins
Topoisomerase

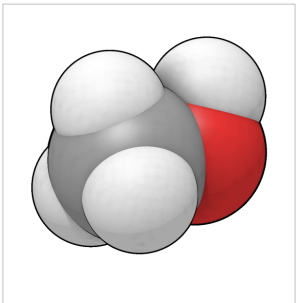


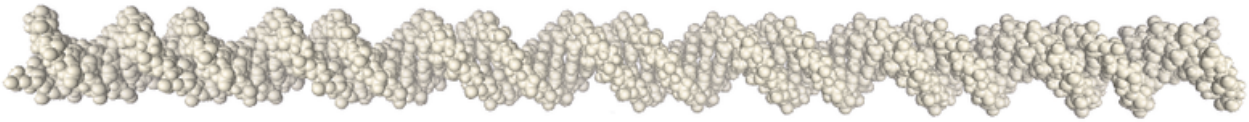
A-DNA B-DNA


A
B
C
D



- Spacefilling model or CPK model - a molecule is represented by overlapping spheres representing the atoms.





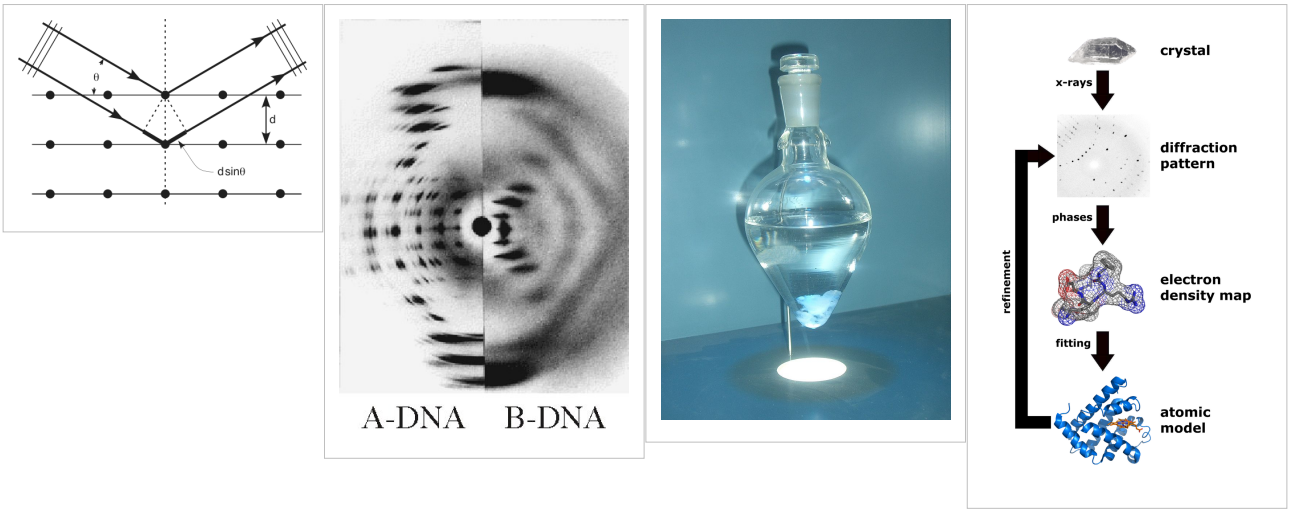
DNA Spacefilling molecular model

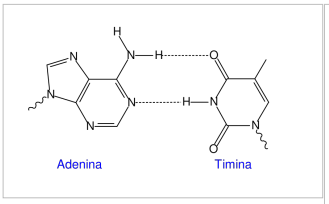
Images for DNA Structure Determination from X-Ray Patterns

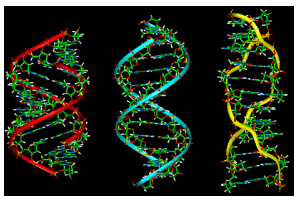
The following images illustrate both the principles and the main steps involved in generating structural information from X-ray diffraction studies of oriented DNA fibers with the help of molecular models of DNA that are combined with crystallographic and mathematical analysis of the X-ray patterns. From left to right the gallery of images shows:

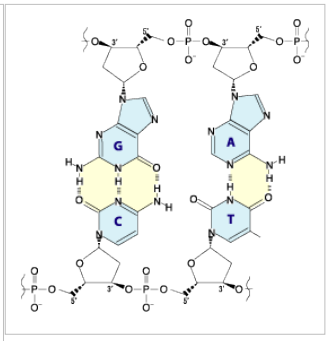
- *First row*:
- 1. Constructive X-ray interference, or diffraction, following Bragg's Law of X-ray "*reflection by the crystal planes*";
- 2. A comparison of A-DNA (crystalline) and highly hydrated B-DNA (paracrystalline) X-ray diffraction, and respectively, X-ray scattering patterns (courtesy of Dr. Herbert R. Wilson, FRS- see refs. list);
- 3. Purified DNA precipitated in a water jug;
- 4. The major steps involved in DNA structure determination by X-ray crystallography showing the important role played by molecular models of DNA structure in this iterative, structure--determination process;
- *Second row*:

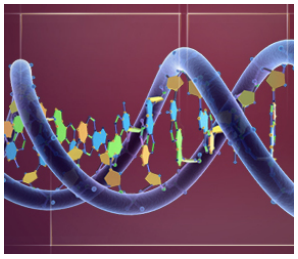
- 5. Photo of a modern X-ray diffractometer employed for recording X-ray patterns of DNA with major components: X-ray source, goniometer, sample holder, X-ray detector and/or plate holder;
- 6. Illustrated animation of an X-ray goniometer;
- 7. X-ray detector at the SLAC synchrotron facility;
- 8. Neutron scattering facility at ISIS in UK;
  - *Third and fourth rows*: *Molecular models of DNA structure at various scales*; figure #11 is an actual electron micrograph of a DNA fiber bundle, presumably of a single bacterial chromosome loop.

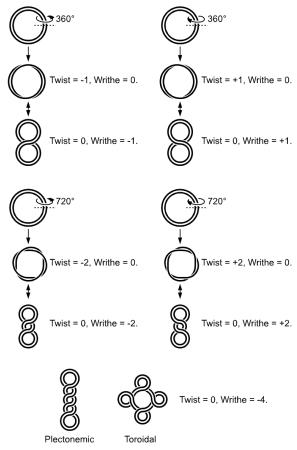


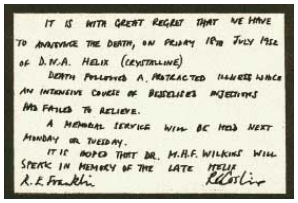






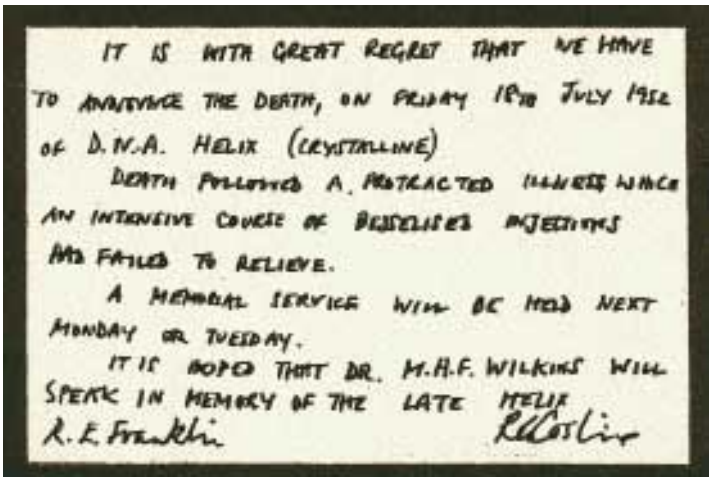






Paracrystalline lattice models of B-DNA structures

A paracrystalline lattice, or paracrystal, is a molecular or atomic lattice with significant amounts (e.g., larger than a few percent) of partial disordering of molecular arranegements. Limiting cases of the paracrystal model are nanostructures, such as glasses, liquids, etc., that may possess only local ordering and no global order. Liquid crystals also have paracrystalline rather than crystalline structures.



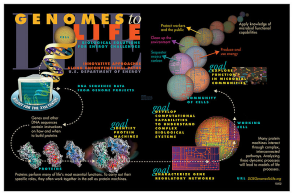
DNA Helix controversy in 1952

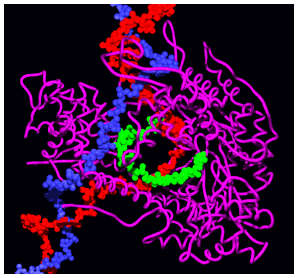
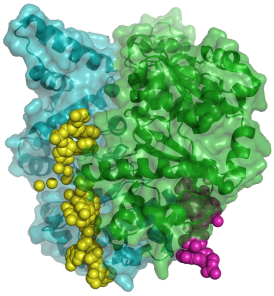
Highly hydrated B-DNA occurs naturally in living cells in such a paracrystalline state, which is a dynamic one in spite of the relatively rigid DNA double-helix stabilized by parallel hydrogen bonds between the nucleotide base-pairs in the two complementary, helical DNA chains (see figures). For simplicity most DNA molecular models ommit both water and ions dynamically bound to B-DNA, and are thus less useful for understanding the dynamic behaviors of B-DNA *in vivo*. The physical and mathematical analysis of X-ray[16] [17] and spectroscopic data for paracrystalline B-DNA is therefore much more complicated than that of crystalline, A-DNA X-ray diffraction patterns. The paracrystal model is also important for DNA technological applications such as DNA nanotechnology. Novel techniques that combine X-ray diffraction of DNA with X-ray microscopy in hydrated living cells are now also being developed (see, for example, "Application of X-ray microscopy in the analysis of living hydrated cells" [18]).

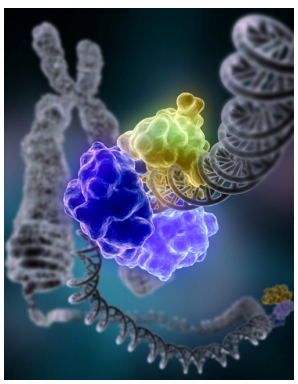
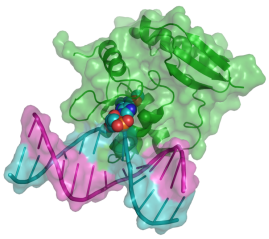
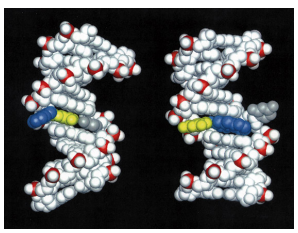
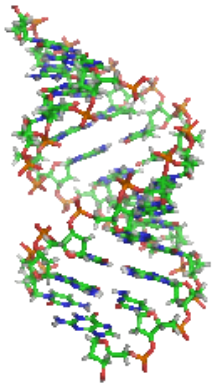
Genomic and Biotechnology Applications of DNA molecular modeling

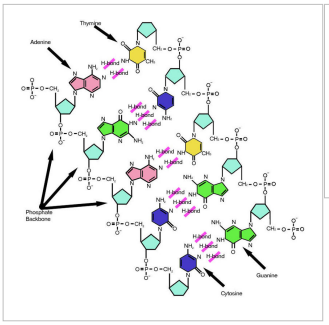
The following gallery of images illustrates various uses of DNA molecular modeling in Genomics and Biotechnology research applications from DNA repair to PCR and DNA nanostructures; each slide contains its own explanation and/or details. The first slide presents an overview of DNA applications, including DNA molecular models, with emphasis on Genomics and Biotechnology.

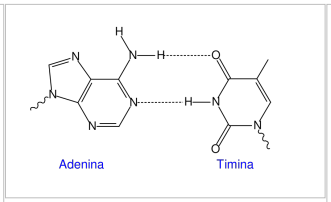
Gallery: DNA Molecular modeling applications

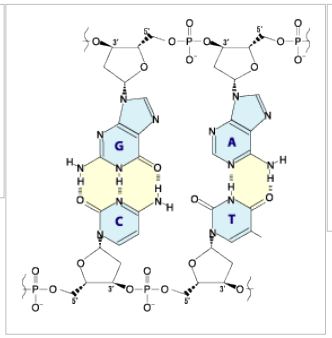


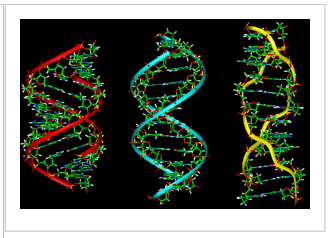



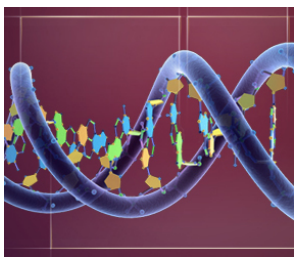





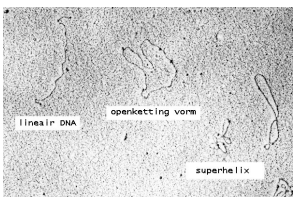

Thymine
Adenine
Phosphate Backbone
Guanine
Cytosine

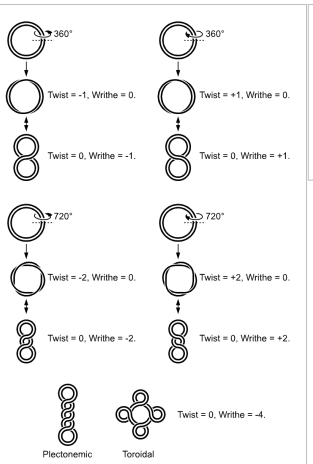

Adenina
Timina

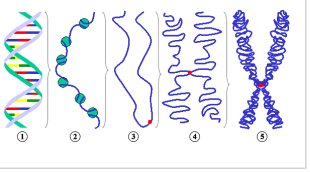

G
C
A
T

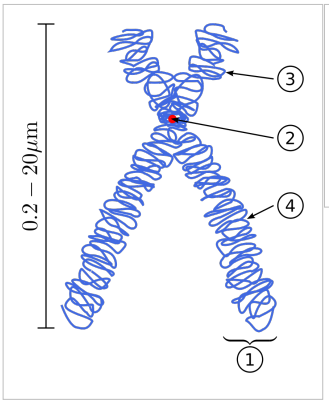


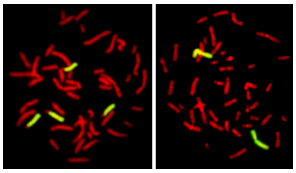


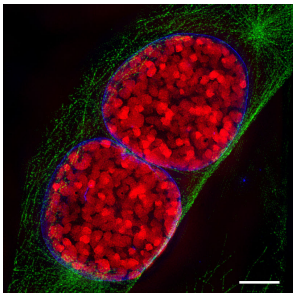

linear DNA
openkettig vorm
superhelix

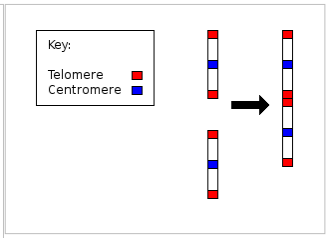

360°
Twist = -1, Writhe = 0.
Twist = 0, Writhe = -1.
360°
Twist = +1, Writhe = 0.
Twist = 0, Writhe = +1.
720°
Twist = -2, Writhe = 0.
Twist = 0, Writhe = -2.
720°
Twist = +2, Writhe = 0.
Twist = 0, Writhe = +2.
Twist = 0, Writhe = -4.
Plectonemic
Toroidal

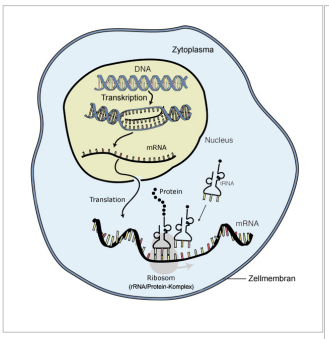

1
2
3
4
5

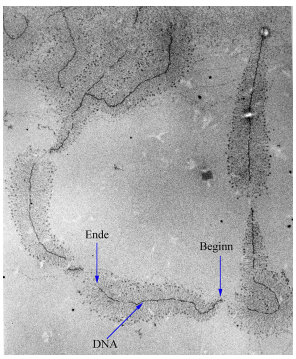

0.2 – 20µm
1
2
3
4

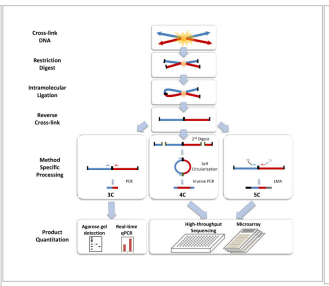


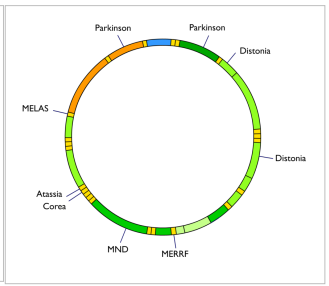


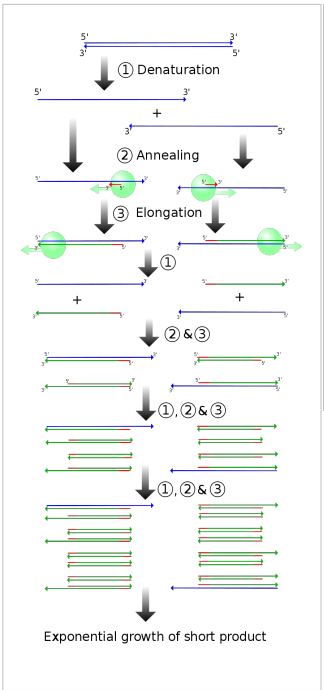

Key:
Telomere
Centromere

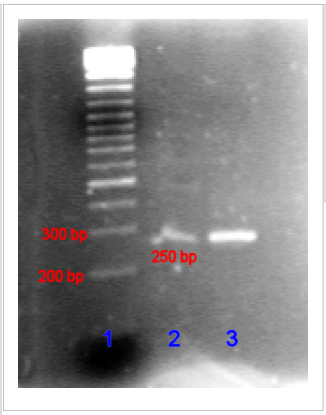

Zytoplasma
DNA
Transkription
Nucleus
mRNA
Translation
Protein
tRNA
mRNA
Ribosom
(rRNA/Protein-Komplex)
Zellmembran

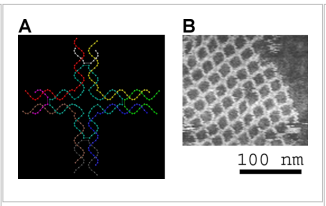

Ende
Beginn
DNA

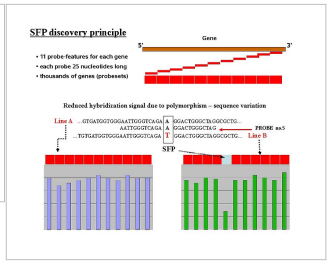

Cross-link DNA
Restriction Digest
Intramolecular Ligation
Reverse Cross-link
Method Specific Processing
3C
4C
5C
Product Quantitation
Agarose gel detection
Real-time qPCR
High-throughput sequencing
Microarray

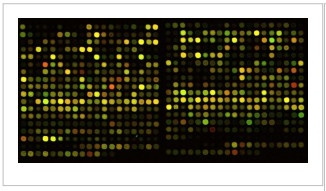

Parkinson
Parkinson
Distonia
MELAS
Distonia
Atassia
Corea
MND
MERRF

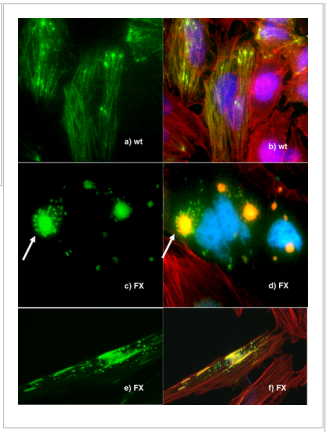


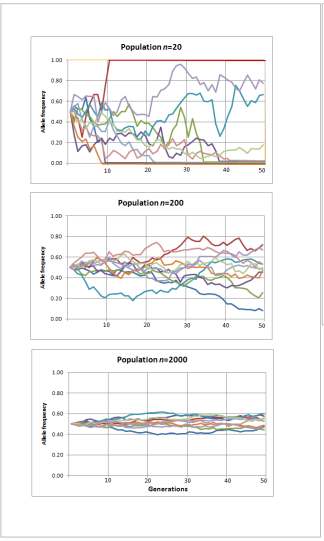












Databases for DNA molecular models and sequences

X-ray diffraction

- NDB ID: UD0017 Database [13]
- X-ray Atlas -database [19]
- PDB files of coordinates for nucleic acid structures from X-ray diffraction by NA (incl. DNA) crystals [20]
- Structure factors dowloadable files in CIF format [21]

Neutron scattering

- ISIS neutron source
- ISIS pulsed neutron source:A world centre for science with neutrons & muons at Harwell, near Oxford, UK. [22]

X-ray microscopy

- Application of X-ray microscopy in the analysis of living hydrated cells [18]

Electron microscopy

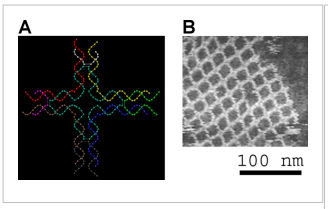
- DNA under electron microscope [23]

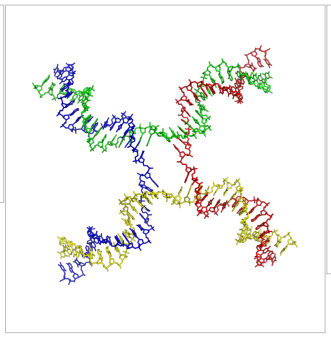
Atomic Force Microscopy (AFM)

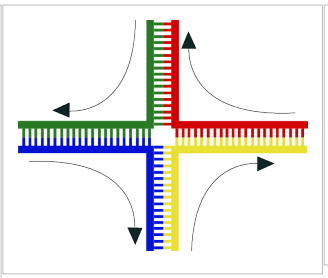
Two-dimensional DNA junction arrays have been visualized by Atomic Force Microscopy (AFM)[24] . Other imaging resources for AFM/Scanning probe microscopy(SPM) can be freely accessed at:

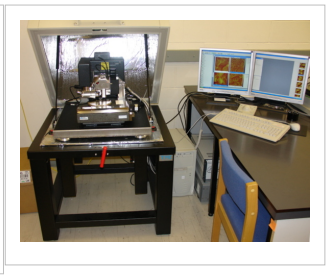
- How SPM Works [25]
- SPM Image Gallery - AFM STM SEM MFM NSOM and more. [26]

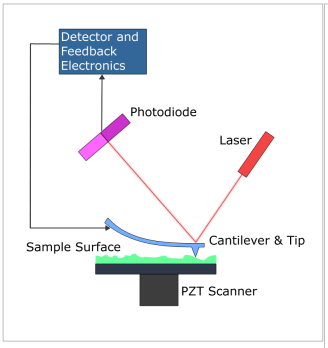
Gallery of AFM Images

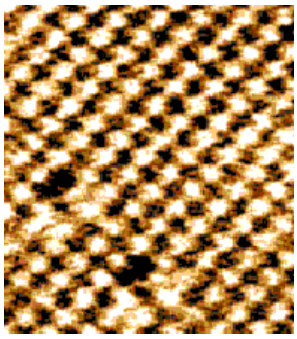


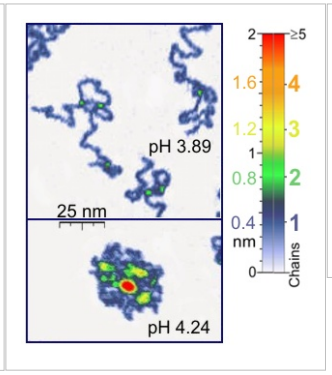


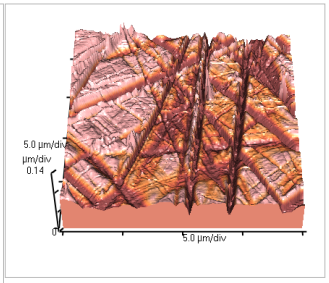




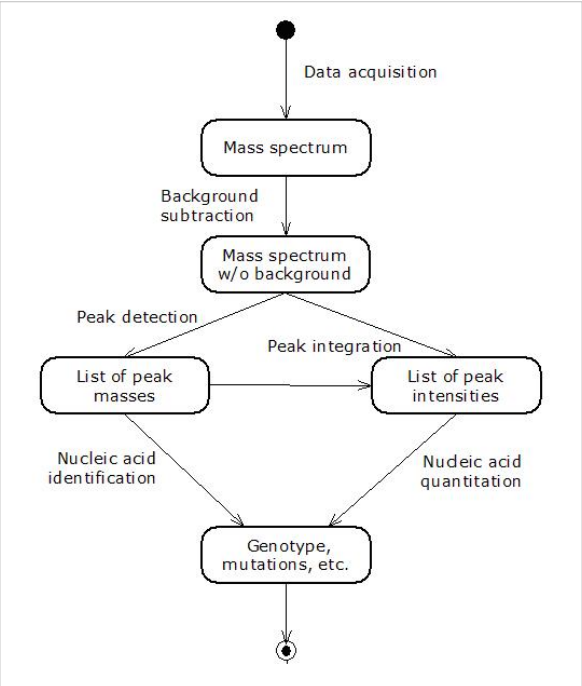








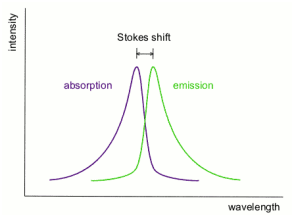
Mass spectrometry--Maldi informatics

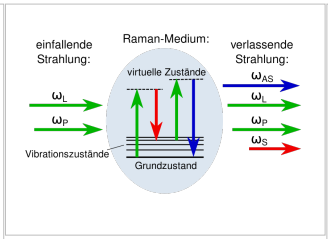


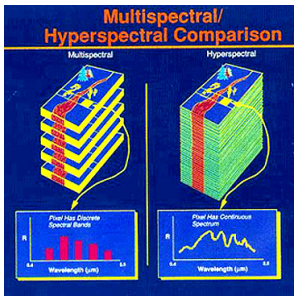
Spectroscopy

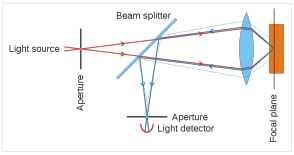
- Vibrational circular dichroism (VCD)
- FT-NMR[27] [28]
  - NMR Atlas--database [29]
  - mmcif downloadable coordinate files of nucleic acids in solution from 2D-FT NMR data [30]
  - NMR constraints files for NAs in PDB format [31]
- NMR microscopy[32]
- Microwave spectroscopy
- FT-IR
- FT-NIR[33] [34] [35]
- Spectral, Hyperspectral, and Chemical imaging)[36] [37] [38] [39] [40] [41] [42] .
- Raman spectroscopy/microscopy[43] and CARS[44] .
- Fluorescence correlation spectroscopy[45] [46] [47] [48] [49] [50] [51] [52] , Fluorescence cross-correlation spectroscopy and FRET[53] [54] [55] .
- Confocal microscopy[56]

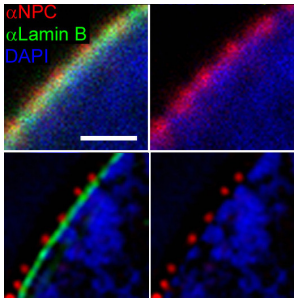
Gallery: CARS (Raman spectroscopy), Fluorescence confocal microscopy, and Hyperspectral imaging

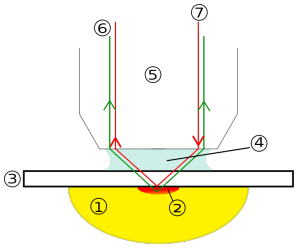


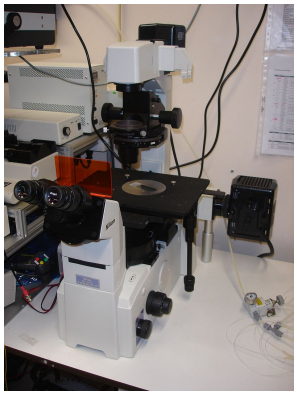


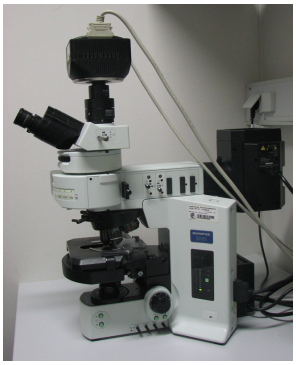


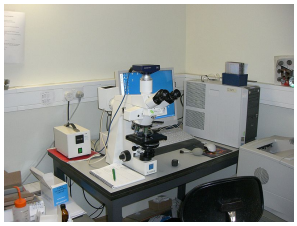


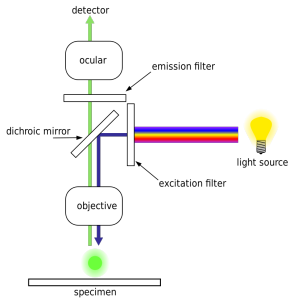


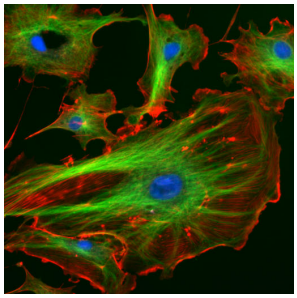


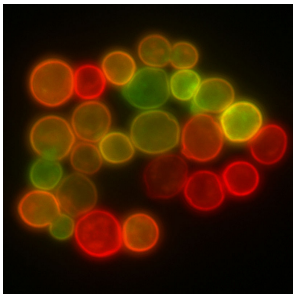


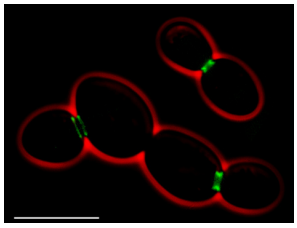


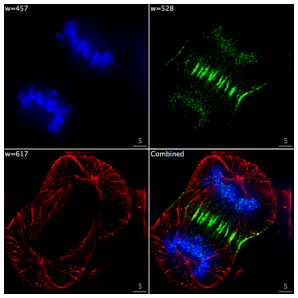


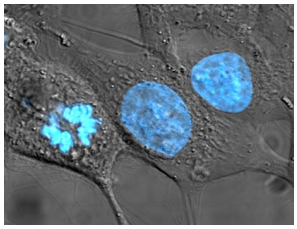


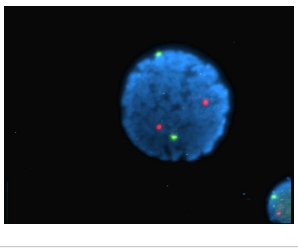


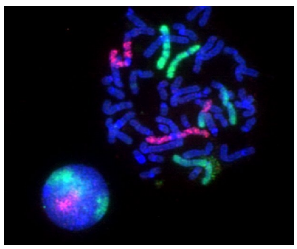


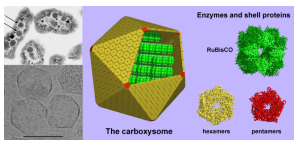












Genomic and structural databases

- CBS Genome Atlas Database [57] — contains examples of base skews.[58]
- The Z curve database of genomes — a 3-dimensional visualization and analysis tool of genomes [59][60] .
- DNA and other nucleic acids' molecular models: Coordinate files of nucleic acids molecular structure models in PDB and CIF formats [61]

Notes

[1] Franklin, R.E. and Gosling, R.G. recd.6 March 1953. Acta Cryst. (1953). 6, 673 The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content Acta Cryst. (1953). and 6, 678 The Structure of Sodium Thymonucleate Fibres II. The Cylindrically Symmetrical Patterson Function.

[2] Cochran, W., Crick, F.H.C. and Vand V. 1952. The Structure of Synthetic Polypeptides. 1. The Transform of Atoms on a Helix. Acta Cryst. 5(5):581-586.

[3] Crick, F.H.C. 1953a. The Fourier Transform of a Coiled-Coil., Acta Crystallographica 6(8-9):685-689.

[4] Watson, J.D; Crick F.H.C. 1953a. Molecular Structure of Nucleic Acids- A Structure for Deoxyribose Nucleic Acid., Nature 171(4356):737-738.

[5] Watson, J.D; Crick F.H.C. 1953b. The Structure of DNA., Cold Spring Harbor Symposia on Qunatitative Biology 18:123-131.

[6] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). "http://www.nature.com/nature/dna50/wilkins.pdf|Molecular Structure of Deoxypentose Nucleic Acids" (PDF). *Nature* **171**: 738–740. doi: 10.1038/171738a0 (http://dx.doi.org/10.1038/171738a0). PMID 13054693. http://www.nature.com/nature/dna50/wilkins.pdf.

[7] Elson D, Chargaff E (1952). "On the deoxyribonucleic acid content of sea urchin gametes". *Experientia* **8** (4): 143-145.

[8] Chargaff E, Lipshitz R, Green C (1952). "Composition of the deoxypentose nucleic acids of four genera of sea-urchin". *J Biol Chem* **195** (1): 155-160. PMID 14938364.

[9] Chargaff E, Lipshitz R, Green C, Hodes ME (1951). "The composition of the deoxyribonucleic acid of salmon sperm". *J Biol Chem* **192** (1): 223-230. PMID 14917668.

[10] Chargaff E (1951). "Some recent studies on the composition and structure of nucleic acids". *J Cell Physiol Suppl* **38** (Suppl).

[11] Magasanik B, Vischer E, Doniger R, Elson D, Chargaff E (1950). "The separation and estimation of ribonucleotides in minute quantities". *J Biol Chem* **186** (1): 37-50. PMID 14778802.

[12] Chargaff E (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation". *Experientia* **6** (6): 201-209.

[13] http://ndbserver.rutgers.edu/atlas/xray/structures/U/ud0017/ud0017.html

[14] http://www.phy.cam.ac.uk/research/bss/molbiophysics.php

[15] http://planetphysics.org/encyclopedia/TheoreticalBiophysics.html

[16] Hosemann R., Bagchi R.N., *Direct analysis of diffraction by matter*, North-Holland Publs., Amsterdam – New York, 1962.

[17] Baianu, I.C. (1978). "X-ray scattering by partially disordered membrane systems.". *Acta Cryst.,* **A34** (5): 751–753. doi: 10.1107/S0567739478001540 (http://dx.doi.org/10.1107/S0567739478001540).

[18] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938

[19] http://ndbserver.rutgers.edu/atlas/xray/index.html

[20] http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-biol/

[21] http://ndbserver.rutgers.edu/ftp/NDB/structure-factors/
[22] http://www.isis.rl.ac.uk/
[23] http://www.fidelitysystems.com/Unlinked_DNA.html
[24] Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
[25] http://www.parkafm.com/New_html/resources/01general.php
[26] http://www.rhk-tech.com/results/showcase.php
[27] (http://www.jonathanpmiller.com/Karplus.html)- obtaining dihedral angles from ^{3}J coupling constants
[28] (http://www.spectroscopynow.com/FCKeditor/UserFiles/File/specNOW/HTML files/General_Karplus_Calculator.htm) Another Javascript-like NMR coupling constant to dihedral
[29] http://ndbserver.rutgers.edu/atlas/nmr/index.html
[30] http://ndbserver.rutgers.edu/ftp/NDB/coordinates/na-nmr-mmcif/
[31] http://ndbserver.rutgers.edu/ftp/NDB/nmr-restraints/
[32] Lee, S. C. et al., (2001). One Micrometer Resolution NMR Microscopy. *J. Magn. Res.*, **150**: 207-213.
[33] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[34] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[35] Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
[36] http://www.imaging.net/chemical-imaging/Chemical imaging
[37] http://www.malvern.com/LabEng/products/sdi/bibliography/sdi_bibliography.htm E. N. Lewis, E. Lee and L. H. Kidder, Combining Imaging and Spectroscopy: Solving Problems with Near-Infrared Chemical Imaging. Microscopy Today, Volume 12, No. 6, 11/2004.
[38] D.S. Mantus and G. H. Morrison. 1991. Chemical imaging in biology and medicine using ion microscopy., *Microchimica Acta*, **104**, (1-6) January 1991, doi: 10.1007/BF01245536
[39] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[40] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[41] J. Dubois, G. Sando, E. N. Lewis, Near-Infrared Chemical Imaging, A Valuable Tool for the Pharmaceutical Industry, G.I.T. Laboratory Journal Europe, No.1-2, 2007.
[42] Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology.(June 2004).,I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin q-bio/0406047 (http://arxiv.org/abs/q-bio/0406047)
[43] Chemical Imaging Without Dyeing (http://witec.de/en/download/Raman/ImagingMicroscopy04.pdf)
[44] C.L. Evans and X.S. Xie.2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine., doi:10.1146/annurev.anchem.1.031207.112754 *Annual Review of Analytical Chemistry*, **1**: 883-909.
[45] Eigen, M., Rigler, M. Sorting single molecules: application to diagnostics and evolutionary biotechnology,(1994) *Proc. Natl. Acad. Sci. USA*, 91,5740-5747.
[46] Rigler, M. Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology,(1995) J. Biotechnol., 41,177-186.
[47] Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
[48] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[49] Oehlenschläger F., Schwille P. and Eigen M. (1996). Detection of HIV-1 RNA by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. USA* **93**:1281.
[50] Bagatolli, L.A., and Gratton, E. (2000). Two-photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. Biophys J., 78:290-305.

[51] Schwille, P., Haupts, U., Maiti, S., and Webb. W.(1999). Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. Biophysical Journal, 77(10):2251-2265.
[52] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[53] FRET description (http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/fret.htm)
[54] doi:10.1016/S0959-440X(00)00190-1 (http://dx.doi.org/10.1016/S0959-440X(00)00190-1)Recent advances in FRET: distance determination in protein–DNA complexes. *Current Opinion in Structural Biology* **2001**, *11(2)*, 201-207
[55] http://www.fretimaging.org/mcnamaraintro.html FRET imaging introduction
[56] Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, Proc. Natl. Acad. Sci. USA 91:5740.
[57] http://www.cbs.dtu.dk/services/GenomeAtlas/
[58] Hallin PF, David Ussery D (2004). "CBS Genome Atlas Database: A dynamic storage for bioinformatic results and DNA sequence data". *Bioinformatics* **20**: 3682-3686.
[59] http://tubic.tju.edu.cn/zcurve/
[60] Zhang CT, Zhang R, Ou HY (2003). "The Z curve database: a graphic representation of genome sequences". Bioinformatics 19 (5): 593-599. doi:10.1093/bioinformatics/btg041
[61] http://ndbserver.rutgers.edu/ftp/NDB/models/

References

- *Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology*.(June 2004) I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin., q-bio/0406047.
- F. Bessel, *Untersuchung des Theils der planetarischen Störungen*, Berlin Abhandlungen (1824), article 14.
- Sir Lawrence Bragg, FRS. *The Crystalline State, A General survey.* London: G. Bells and Sons, Ltd., vols. 1 and 2., 1966., 2024 pages.
- Cantor, C. R. and Schimmel, P.R. *Biophysical Chemistry, Parts I and II.*, San Franscisco: W.H. Freeman and Co. 1980. 1,800 pages.
- Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, *Proc. Natl. Acad. Sci. USA* **91**:5740.
- Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
- Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
- Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004. I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004).
- Voet, D. and J.G. Voet. *Biochemistry*, 2nd Edn., New York, Toronto, Singapore: John Wiley & Sons, Inc., 1995, ISBN: 0-471-58651-X., 1361 pages.
- Watson, G. N. *A Treatise on the Theory of Bessel Functions.*, (1995) Cambridge University Press. ISBN 0-521-48391-3.
- Watson, James D. and Francis H.C. Crick. A structure for Deoxyribose Nucleic Acid (http://www.nature.com/nature/dna50/watsoncrick.pdf) (PDF). *Nature* 171, 737–738, 25 April 1953.
- Watson, James D. *Molecular Biology of the Gene.* New York and Amsterdam: W.A. Benjamin, Inc. 1965., 494 pages.

- Wentworth, W.E. *Physical Chemistry. A short course.*, Malden (Mass.): Blackwell Science, Inc. 2000.
- Herbert R. Wilson, FRS. *Diffraction of X-rays by proteins, Nucleic Acids and Viruses.*, London: Edward Arnold (Publishers) Ltd. 1966.
- Kurt Wuthrich. *NMR of Proteins and Nucleic Acids.*, New York, Brisbane,Chicester, Toronto, Singapore: J. Wiley & Sons. 1986., 292 pages.
- Robinson, Bruche H.; Seeman, Nadrian C. (August 1987). "The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device". *Protein Engineering* **1** (4): 295–300. ISSN 0269-2139 (http://worldcat.org/issn/0269-2139). Link (http://peds.oxfordjournals.org/cgi/content/abstract/1/4/295)
- Rothemund, Paul W. K.; Ekani-Nkodo, Axel; Papadakis, Nick; Kumar, Ashish; Fygenson, Deborah Kuchnir & Winfree, Erik (22 December 2004). "Design and Characterization of Programmable DNA Nanotubes". *Journal of the American Chemical Society* **126** (50): 16344–16352. doi: 10.1021/ja044319l (http://dx.doi.org/10.1021/ja044319l). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Keren, K.; Kinneret Keren, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, Erez Braun (November 2003). "http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380|DNA-Templated Carbon Nanotube Field-Effect Transistor". *Science* **302** (6549): 1380–1382. doi: 10.1126/science.1091022 (http://dx.doi.org/10.1126/science.1091022). ISSN 1095-9203 (http://worldcat.org/issn/1095-9203). http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380.
- Zheng, Jiwen; Constantinou, Pamela E.; Micheel, Christine; Alivisatos, A. Paul; Kiehl, Richard A. & Seeman Nadrian C. (2006). "2D Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs". *Nano Letters* **6**: 1502–1504. doi: 10.1021/nl060994c (http://dx.doi.org/10.1021/nl060994c). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Cohen, Justin D.; Sadowski, John P.; Dervan, Peter B. (2007). "Addressing Single Molecules on DNA Nanostructures". *Angewandte Chemie* **46** (42): 7956–7959. doi: 10.1002/anie.200702767 (http://dx.doi.org/10.1002/anie.200702767). ISSN 0570-0833 (http://worldcat.org/issn/0570-0833).
- Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Constantinou, Pamela E.; Wang, Tong; Kopatsch, Jens; Israel, Lisa B.; Zhang, Xiaoping; Ding, Baoquan; Sherman, William B.; Wang, Xing; Zheng, Jianping; Sha, Ruojie & Seeman, Nadrian C. (2006). "Double cohesion in structural DNA nanotechnology". *Organic and Biomolecular Chemistry* **4**: 3414–3419. doi: 10.1039/b605212f (http://dx.doi.org/10.1039/b605212f).

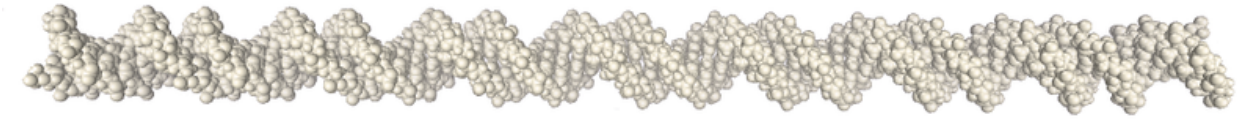
See also

- DNA
- Molecular graphics
- DNA structure
- DNA Dynamics
- X-ray scattering
- Neutron scattering
- Crystallography
- Crystal lattices
- Paracrystalline lattices/Paracrystals
- 2D-FT NMRI and Spectroscopy
- NMR Spectroscopy
- Microwave spectroscopy
- Two-dimensional IR spectroscopy
- Spectral imaging
- Hyperspectral imaging
- Chemical imaging
- NMR microscopy
- VCD or Vibrational circular dichroism
- FRET and FCS- Fluorescence correlation spectroscopy
- Fluorescence cross-correlation spectroscopy (FCCS)
- Molecular structure
- Molecular geometry
- Molecular topology
- DNA topology
- Sirius visualization software
- Nanostructure
- DNA nanotechnology
- Imaging
- Atomic force microscopy
- X-ray microscopy
- Liquid crystal
- Glasses
- QMC@Home
- Sir Lawrence Bragg, FRS
- Sir John Randall
- James Watson
- Francis Crick
- Maurice Wilkins
- Herbert Wilson, FRS
- Alex Stokes

External links

- DNA the Double Helix Game (http://nobelprize.org/educational_games/medicine/dna_double_helix/) From the official Nobel Prize web site
- MDDNA: Structural Bioinformatics of DNA (http://humphry.chem.wesleyan.edu:8080/MDDNA/)
- Double Helix 1953–2003 (http://www.ncbe.reading.ac.uk/DNA50/) National Centre for Biotechnology Education
- DNA under electron microscope (http://www.fidelitysystems.com/Unlinked_DNA.html)
- Ascalaph DNA (http://www.agilemolecule.com/Ascalaph/Ascalaph_DNA.html) — Commercial software for DNA modeling
- DNAlive: a web interface to compute DNA physical properties (http://mmb.pcb.ub.es/DNAlive). Also allows cross-linking of the results with the UCSC Genome browser and DNA dynamics.
- DiProDB: Dinucleotide Property Database (http://diprodb.fli-leibniz.de). The database is designed to collect and analyse thermodynamic, structural and other dinucleotide properties.
- Further details of mathematical and molecular analysis of DNA structure based on X-ray data (http://planetphysics.org/encyclopedia/BesselFunctionsApplicationsToDiffractionByHelicalStructures.html)
- Bessel functions corresponding to Fourier transforms of atomic or molecular helices. (http://planetphysics.org/?op=getobj&from=objects&name=BesselFunctionsAndTheirApplicationsToDiffractionByHelicalStructures)
- Application of X-ray microscopy in analysis of living hydrated cells (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938)
- Characterization in nanotechnology some pdfs (http://nanocharacterization.sitesled.com/)
- overview of STM/AFM/SNOM principles with educative videos (http://www.ntmdt.ru/SPM-Techniques/Principles/)
- SPM Image Gallery - AFM STM SEM MFM NSOM and More (http://www.rhk-tech.com/results/showcase.php)
- How SPM Works (http://www.parkafm.com/New_html/resources/01general.php)
- U.S. National DNA Day (http://www.genome.gov/10506367) — watch videos and participate in real-time discusssions with scientists.
- The Secret Life of DNA - DNA Music compositions (http://www.tjmitchell.com/stuart/dna.html)

List of nucleic acid simulation software



This is a list of computer programs that are used for nucleic acids simulations.

Min - Optimization, **MD** - Molecular Dynamics, **MC** - Monte Carlo,

Crt - Cartesian coordinates. **Int** - Internal coordinates **Exp** - Explicit water. **Imp** - Implicit water.

Lig - Ligands interactions. **HA** - Hardware accelerated.

Name	View 3D	Model Build	Min	MD	MC	Crt	Int	Exp	Imp	Lig	HA	Comments	License	Homepage
Abalone	+	+	+	+		+		+	+	+		DNA, proteins, ligands	Commercial	Agile Molecule [1]
AMBER [1]		+	+	+		+		+	+	+		AMBER Force Field	Commercial	ambermd.org [5]
CHARMM		+	+	+	+	+		+	+	+		CHARMM Force Field	Commercial	charmm.org [9]
ICM [2]	+	+	+		+		+		+			Global optimization	Commercial	Molsoft [3]
JUMNA [4]		+	+				+		+				Commercial	
MDynaMix [5]	+	+		+		+		+		+		Common MD	GPL	Stockholm University [21]
MOE	+	+	+	+		+		+		+		**M**olecular **O**perating **E**nvironment	Commercial	Chemical Computing Group [22]
NAB [6]		+										**N**ucleic **A**cid **B**uilder	GPL	New Jersey University [7]
NAMD	+		+	+		+		+		+	+	**NA**noscale **M**olecular **D**ynamics	Free	Illinois University [8]

See also

- Molecular Modelling
- Molecular graphics
- Molecular mechanics
- Molecular dynamics
- Molecular Design software
- Quantum chemistry computer programs
- List of RNA structure prediction software
- List of protein structure prediction software
- List of software for molecular mechanics modeling
- Force field
- Force field implementation

References

[1] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Jr., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W. and Kollman P.A. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *J. Am. Chem. Soc.* **117**: 5179-5197.

[2] Abagyan R.A., Totrov M.M. and Kuznetsov D.A. (1994). "Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation". *J. Comp. Chem.* **15**: 488-506.

[3] http://www.molsoft.com

[4] Lavery, R., Zakrzewska, K. and Sklenar, H. (1995). "JUMNA: junction minimisation of nucleic acids". *Comp. Phys. Commun.* **91**: 135-158.

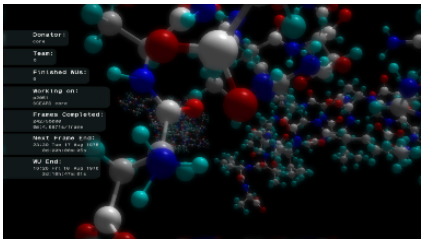
[5] A.P.Lyubartsev, A.Laaksonen (2000). "MDynaMix - A scalable portable parallel MD simulation package for arbitrary molecular mixtures". *Computer Physics Communications* **128**: 565-589.

[6] Macke T. and Case D.A. (1998). "Modeling unusual nucleic acid structures". *Molecular Modeling of Nucleic Acids*: 379-393.

[7] http://casegroup.rutgers.edu

[8] http://www.ks.uiuc.edu/Research/namd/

Folding@home



The PlayStation 3 Folding@home client displays a 3D model of the protein being simulated

Original author(s)	Vijay Pande
Developer(s)	Stanford University / Pande Group
Initial release	2000-10-01
Stable release	*Windows:* 6.23 (Uniprocessor) 6.23 (GPU) *Mac OS X:* 6.20 (PPC-Uniprocessor) 6.20 (x86-SMP) *Linux:* 6.02 (Uniprocessor) 6.02 (x64-SMP) *PlayStation 3:* 1.4 [1] / 2008-11-26 (Windows 6.23)
Preview release	6.23beta (Windows SMP) 6.24beta (Linux x64-SMP) 6.24beta (Mac OS X x86-SMP) / 2009-01-20 (6.24betas)
Platform	Cross-platform
Available in	English
Type	Distributed computing
License	Proprietary [2]
Website	folding.stanford.edu [3]

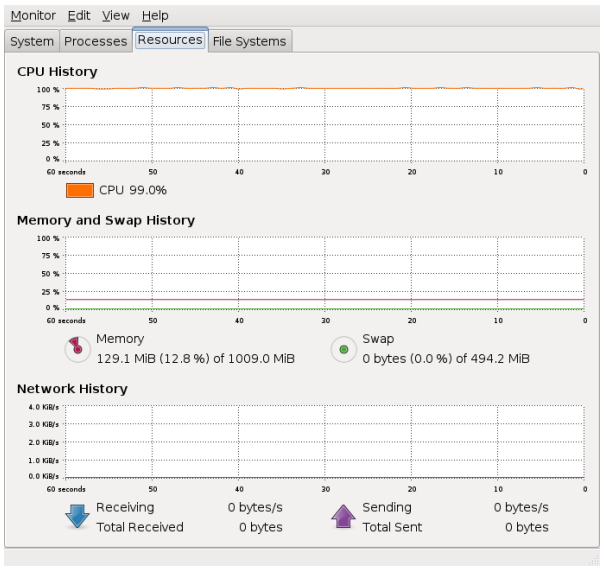
Folding@home (sometimes abbreviated as **FAH** or **F@h**) is a distributed computing (DC) project designed to perform computationally intensive simulations of protein folding and other molecular dynamics (MD). It was launched on October 1, 2000, and is currently managed by the Pande Group, within Stanford University's chemistry department, under the supervision of Professor Vijay Pande. Folding@home is the most powerful distributed computing cluster in the world, according to Guinness,[4] and one of the world's largest distributed computing projects.[5] The goal of the project is "to understand protein folding, misfolding, and related diseases."[6]

Purpose

Accurate simulations of protein folding and misfolding enable the scientific community to better understand the development of many diseases, including sickle-cell disease (drepanocytosis), Alzheimer's disease, Parkinson's disease, mad cow disease, cancer, Huntington's disease, cystic fibrosis, osteogenesis imperfecta, alpha 1-antitrypsin deficiency, and other aggregation-related diseases.[7] More fundamentally, understanding the process of protein folding — how biological molecules assemble themselves into a functional state — is one of the outstanding problems of molecular biology. So far, the Folding@home project has successfully simulated folding in the 5-10 microsecond range — which is a far longer simulation than it was previously thought possible to model. The Pande Group goal is to refine and improve the MD and Folding@home DC methods to the level where it will become an essential tool for the MD research.[8] For that goal they collaborate with various scientific institutions.[9] As of February 19, 2009, sixty-three scientific research papers have been published using the project's work.[10] A University of Illinois at Urbana-Champaign report dated October 22, 2002 states that Folding@home distributed simulations of protein folding are demonstrably accurate.[11]

Function

Folding@home does not rely on powerful supercomputers for its data processing; instead, the primary contributors to the Folding@home project are many hundreds of thousands of personal computer users who have installed a small client program. The client will, at the user's choice, run in the background, utilizing otherwise unused CPU power, or run as a screensaver only while the user is away. In most modern personal computers, the CPU is rarely used to its full capacity at all times; the Folding@home client takes advantage of this unused processing power.



Folding@home when running takes advantage of unused CPU cycles on a computer system as shown by this computer's 99% CPU usage.

The Folding@home client periodically connects to a server to retrieve "work units", which are packets of data upon which to perform calculations. Each completed work unit is then sent back to the server. As data integrity is a major concern for all distributed computing projects, all work units are validated through the use of a 2048 bit digital signature.

Contributors to Folding@home may have user names used to keep track of their contributions. Each user may be running the client on one or more CPUs; for example, a user with two computers could run the client on both of them. Users may also contribute under one or more team names; many different users may join together to form a team. Contributors are assigned a score indicating the number and difficulty of completed work units. Rankings and other statistics are posted to the Folding@home website.

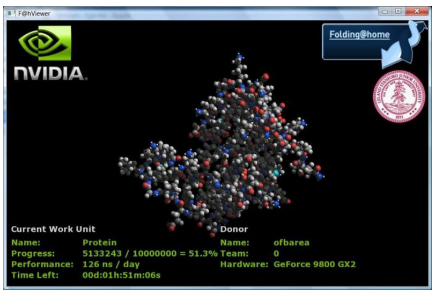
Analysis Software

The Folding@home client utilizes modified versions of five molecular simulation programs for calculation: TINKER, GROMACS, AMBER, CPMD, and SHARPEN.[12] Where possible, optimizations are used to speed the process of calculation. There are many variations on these base simulation programs, each of which is given an arbitrary identifier (Core xx):[13]

Active Cores

- GROMACS (all variants of this core use SIMD optimizations including SSE, 3DNow+ or AltiVec, where available, unless otherwise specified)
  - Gromacs (Core 78)
    - Available for all Uniprocessor clients only.
  - DGromacs (Core 79)
    - Double precision Gromacs, uses SSE2 only.
    - Available for all Uniprocessor clients only.
  - DGromacsB (Core 7b)
    - Nominally an update of DGromacs, but is actually based on the SMP/GPU codebases (and is therefore a completely new core). As a result, both are still in use.
    - Double precision Gromacs, uses SSE2 only.
    - Available for all Uniprocessor clients only.
  - DGromacsC (Core 7c)
    - Double precision Gromacs, uses SSE2 only.
    - Available on Windows and Linux Uniprocessor clients only.
  - GBGromacs (Core 7a)
    - Gromacs with the Generalized Born implicit solvent model.
    - Available for all Uniprocessor clients only.
  - Gromacs SREM (Core 80)
    - Gromacs Serial Replica Exchange Method.
    - The Gromacs Serial Replica Exchange Method core, also known as GroST (Gromacs Serial replica exchange with Temperatures), uses the Replica Exchange method (also known as REMD or Replica Exchange Molecular Dynamics) in its simulations.
    - Available for Windows and Linux Uniprocessor clients only.
  - GroSimT (Core 81)
    - Gromacs with Simulated Tempering.
    - Available for Windows and Linux Uniprocessor clients only.
  - Gromacs 33 (Core a0)
    - Uses the Gromacs 3.3 codebase.
    - Available for all Uniprocessor clients only.
  - Gro-SMP (Core a1)
    - Symmetric MultiProcessing variant, locked to four threads (but can be run on dual core processors).
    - Runs only on multi-core x86 or x64 hardware, uses SSE only.
    - Available for all SMP clients only.
  - GroCVS (Core a2)
    - Symmetric MultiProcessing variant with scalable numbers of threads.

    - Runs only on multi-core x86 or x64 hardware, with four or more cores, uses SSE only.
    - Uses the Gromacs 4.0 codebase.
    - Available for Linux and Mac OS X SMP clients only.
- GroGPU2 (Core 11)
    - Graphics Processing Unit variant for ATI CAL-enabled and nVidia CUDA-enabled GPUs.
    - Comes in two separate versions, one each for ATI and nVidia, but both have the same Core ID.
    - GPUs do not support SIMD optimizations by design, so none are used in this core.
    - Available for GPU2 client only.
- ATI-DEV (Core 12)



NVIDIA GPU v2.0 r1 client for Windows..

    - Graphics Processing Unit developmental core for ATI CAL-enabled GPUS.
    - Does not support SIMD optimizations.
    - Available for GPU2 client only.
- NVIDIA-DEV (Core 13)
    - Graphics Processing Unit developmental core for nVidia CUDA-enabled GPUs.
    - Does not support SIMD optimizations.
    - Available for GPU2 client only.
- GroGPU2-MT (Core 14)[14]
    - Graphics Processing Unit variant for nVidia CUDA-enabled GPUs.
    - Contains additional debugging code compared to the standard Core 11.
    - Does not support SIMD optimizations.
    - Released March 2, 2009.
    - Available for GPU2 client only.
- Gro-PS3 (Does not have a known ID number, but also called SCEARD core)
    - PlayStation 3 variant.
    - No SIMD optimizations, uses SPE cores for optimization.
    - Available for PS3 client only.

- AMBER
    - PMD (Core 82)[13]
        - No optimizations.
        - Available for Windows and Linux Uniprocessor clients only.

Inactive Cores

- TINKER
  - Tinker core (Core 65)
    - Currently inactive, as the GBGromacs core (Core 7a) performs the same tasks much faster.
    - No optimizations.
    - Available for all Uniprocessor clients only.
- GROMACS
  - GroGPU (Core 10)
    - Graphics Processing Unit variant for ATI series 1xxx GPUs.
    - GPUs do not have optimizations; no SIMD optimizations needed since GPU cores are explicitly designed for SIMD.
    - Inactive as of June 6, 2008 due to end of distribution of GPU1 client units.
    - Available for GPU1 client only.
- CPMD
  - QMD (Core 96)
    - Currently inactive, due to QMD developer graduating from Stanford University and due to current research shifting away from Quantum MD.
    - Caused controversy due to SSE2 issues involving Intel libraries and AMD processors.[15]
    - Uses SSE2 (currently only on Intel CPUs, see above).
    - Available for Windows and Linux Uniprocessor clients only.
- SHARPEN [16]
  - SHARPEN Core [17]
    - Currently inactive, in closed beta testing before general release.
    - Uses different format to standard F@H cores, as there is more than one "Work Unit" (using the normal definition) in each work packet sent to clients.

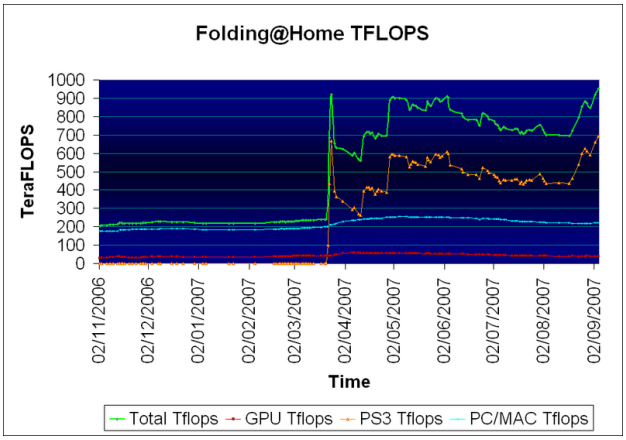
Possible future additions

- ProtoMol [9]

Participation

Shortly after breaking the 200,000 active CPU count on September 20, 2005, the Folding@home project celebrated its fifth anniversary on October 1, 2005.

Interest and participation in the project has grown steadily since its launch. The number of active devices participating in the project increased substantially after receiving much publicity during the launch of their high performance clients for both ATi graphics cards and the PlayStation 3, and again following the launch of the high performance client for nVidia graphics cards.



Folding@home computing power shown - by device type - in TeraFLOPS as recorded semi-daily from November 2006 until September 2007. Note the large spike in total compute power after March 22, when the PlayStation 3 client was released.

As of April 9, 2009 the peak speed of the project overall has reached over 4.5 native PFLOPS (8.1 x86 PFLOPS[18]) from around 400,000 active machines, and the project has received computational results from over 3.75 million devices since it first started.[5]

Google & Folding@home

There used to be cooperation between Folding@home and Google Labs in the form of Google Toolbar. Google Compute supported Folding@home during its early stage — when Folding@home had ~10,000 active CPUs. At that time, a boost of 20,000 machines was very significant. Today the project has a large number of active CPUs and the number of new clients joining Google Compute was very low (most people opted for the Folding@home client instead), so it was discontinued. The Google Compute clients also had certain limits: they could only run the TINKER core and had limited naming and team options. Folding@home is no longer supported on Google Toolbar, and even the old Google Toolbar client will not work.[19]

Genome@home

Folding@home absorbed the Genome@home project on March 8, 2004. The work which was started by the Genome@home project has since been completed using the Folding@home network (the work units without deadlines), and no new work is being distributed by this project. All donators were encouraged to download the Folding@home client (the F@h 4.xx client had a Genome@home option), and once the Genome@home work was complete these clients were asked to donate their processing power to the Folding@home project instead.

PetaFLOPS Milestones

Native petaFLOPS Barrier	Date Crossed
1.0	September 16, 2007
2.0	early May 2008
3.0	August 20, 2008
4.0	September 28, 2008
5.0	February 18, 2009

On September 16, 2007, the Folding@home project officially attained a sustained performance level higher than one native petaFLOPS, becoming the first computing system of any kind in the world to ever do so, although it had briefly peaked above one native petaFLOPS in March 2007, receiving a large amount of main stream media coverage for doing so.[20] [21] In early May 2008 the project attained a sustained performance level higher than two native petaFLOPS, followed by the three and four native petaFLOPS milestones on August 20 and September 28, 2008 respectively. On February 18, 2009, Folding@home achieved a performance level of 5033 native TFLOPS, thereby becoming the first computing system of any kind to surpass 5 native PFLOPS[22] , just as it was for the other four milestones.

The Folding@home computing cluster currently operates at above 4.5 native petaFLOPS at all times, with a large majority of the performance coming from GPU and PlayStation 3 clients.[5] In comparison to this, the fastest standalone supercomputer (non-distributive computing) in the world (as of November 2008, U.S. Department of Energy Roadrunner) peaks at approximately 1.46 petaFLOPS.[23]

Beginning in April 2009, Folding@Home began reporting performance in both "Native" FLOPS and x86 FLOPS.[5] ("x86" FLOPS reported at a much higher mark than the "Native" FLOPS) A detailed explanation of the difference between the two figures was given in the FLOP section of the Folding@Home FAQ.[24]

Results

These peer-reviewed papers (in chronological order) all use research from the Folding@home project.[10]

2000–2001

- M. R. Shirts and V. S. Pande. (2000). "Screen Savers of the World, Unite!". *Science* **290**: 1903–1904. doi:10.1126/science.290.5498.1903 [25]. PMID 17742054.
- Michael R. Shirts and Vijay S. Pande (2001). "Mathematical Analysis of Coupled Parallel Simulations". *Physical Review Letters* **86** (22): 4983–4987. doi:10.1103/PhysRevLett.86.4983 [26].
- Bojan Zagrovic, Eric J. Sorin and Vijay Pande (2001). "b-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model". *Journal of Molecular Biology* **313**: 151–169. doi:10.1006/jmbi.2001.5033 [27].

2002

- Stefan M. Larson, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande (2002) "Folding@home and Genome@home: Using distributed computing to tackle previously intractable problems in computational biology", Stefan M. Larson, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande. To appear in *Computational Genomics*, Richard Grant, editor, Horizon Press
- Bojan Zagrovic, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande. (2002). "Simulation of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwide distributed Computing". *Journal of Molecular Biology* **323**: 927–937. doi:10.1016/S0022-2836(02)00997-X [28].
- Bojan Zagrovic, Christopher D. Snow, Siraj Khaliq, Michael R. Shirts, and Vijay S. Pande (2002). "Native-like Mean Structure in the Unfolded Ensemble of Small Proteins". *Journal of Molecular Biology* **323**: 153–164. doi:10.1016/S0022-2836(02)00888-4 [29].
- Christopher D. Snow, Bojan Zagrovic, and Vijay S. Pande (2002). "The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations". *Journal of the American Chemical Society* **124**: 14548–14549. doi:10.1021/ja028604l [30].

2003

- Vijay S. Pande, Ian Baker, Jarrod Chapman, Sidney P. Elmer, Siraj Khaliq, Stefan M. Larson, Young Min Rhee, Michael R. Shirts, Christopher D. Snow, Eric J. Sorin, Bojan Zagrovic (2003). "Atomistic protein folding simulations on the submillisecond timescale using worldwide distributed computing". *Biopolymers* **68**: 91–109. doi:10.1002/bip.10219 [31].
- Young Min Rhee & Vijay S. Pande (2003). "Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation". *Biophysical Journal* **84** (2): 775–786.
- Eric J. Sorin, Young Min Rhee, Bradley J. Nakatani & Vijay S. Pande (2003). "Insights Into Nucleic Acid Conformational Dynamics from Massively Parallel Stochastic Simulations". *Biophysical Journal* **85**: 790–803.
- Bojan Zagrovic and Vijay S. Pande (2003). "Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study". *Journal of Computational Chemistry* **24** (12): 1432–1436. doi:10.1002/jcc.10297 [32].
- Michael R. Shirts, Jed W. Pitera, William C. Swope, and Vijay S. Pande (2003). "Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins". *Journal of Chemical Physics* **119** (11): 5740–5761. doi:10.1063/1.1587119 [33].
- Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S Pande (2003). "Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods". *Physical Review Letters* **91** (14). doi:10.1103/PhysRevLett.91.140601 [34].
- Bojan Zagrovic & Vijay S Pande (2003). "Structural correspondence between the alpha-helix and the random-flight chain resolves how unfolded proteins can have native-like properties". *Nature Structural Biology* **10** (11): 955–961. doi:10.1038/nsb995 [35].

2004

- Eric J. Sorin, Bradley J. Nakatani, Young Min Rhee, Guha Jayachandran, V Vishal, & Vijay S Pande (2004). "Does Native State Topology Determine the RNA Folding Mechanism?". *Journal of Molecular Biology* **337**: 789–757. doi:10.1016/j.jmb.2004.02.024 [36].
- Christopher D. Snow, Linlin Qiu, Deguo Du, Feng Gai, Stephen J. Hagen, & Vijay S Pande (2004). "Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy". *Proceedings of the National Academy of Sciences, USA* **101** (12): 4077–4082. doi:10.1073/pnas.0305260101 [37].
- Young Min Rhee, Eric J. Sorin, Guha Jayachandran, Erik Lindahl, & Vijay S Pande (2004). "Simulations of the role of water in the protein-folding mechanism". *Proceedings of the National Academy of Sciences, USA* **101** (17): 6456–6461. doi:10.1073/pnas.0307898101 [38].
- Nina Singhal, Christopher D. Snow, and Vijay S. Pande (2004). "Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin". *Journal of Chemical Physics* **121**: 415–425. doi:10.1063/1.1738647 [39].
- L. T. Chong, C. D. Snow, Y. M. Rhee, and V. S. Pande. (2004). "Dimerization of the p53 Oligomerization Domain: Identification of a Folding Nucleus by Molecular Dynamics Simulations". *Journal of Molecular Biology* **345**: 869–878. doi:10.1016/j.jmb.2004.10.083 [40].

2005

- Eric J. Sorin, Young Min Rhee, and Vijay S. Pande (2005). "Does Water Play a Structural Role in the Folding of Small Nucleic Acids?". *Biophysical Journal* **88**: 2516–2524. doi:10.1529/biophysj.104.055087 [41].
- Eric J. Sorin and Vijay S. Pande (2005). "Exploring the Helix-Coil Transition via All-atom Equilibrium Ensemble Simulations". *Biophysical Journal* **88**: 2472–2493. doi:10.1529/biophysj.104.051938 [42].
- Eric J. Sorin and Vijay S. Pande (2005). "Empirical Force-Field Assessment: The Interplay Between Backbone Torsions and Noncovalent Term Scaling". *Journal of Computational Chemistry* **26**: 682–690. doi:10.1002/jcc.20208 [43].
- C. D. Snow, E. J. Sorin, Y. M. Rhee, and V. S. Pande. (2005). "How well can simulation predict protein folding kinetics and thermodynamics?". *Annual Reviews of Biophysics* **34**: 43–69. doi:10.1146/annurev.biophys.34.040204.144447 [44].
- Bojan Zagrovic, Jan Lipfert, Eric J. Sorin, Ian S. Millett, Wilfred F. van Gunsteren, Sebastian Doniach & Vijay S. Pande (2005). "Unusual compactness of a polyproline type II structure". *Proceedings of the National Academy of Sciences, USA* **102** (33): 11698–11703. doi:10.1073/pnas.0409693102 [45].
- Michael R. Shirts & Vijay S. Pande (2005). "Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration". *Journal of Chemical Physics* **122**. doi:10.1063/1.1873592 [46].
- Michael R. Shirts & Vijay S. Pande (2005). "Solvation free energies of amino acid side chain analogs for common molecular mechanics water models". *Journal of Chemical Physics* **122**. doi:10.1063/1.1877132 [47].
- Sidney Elmer, Sanghyun Park, & Vijay S. Pande (2005). "Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular-dynamics simulations in

acetonitrile, chloroform, methanol, and water". *Journal of Chemical Physics* **123**. doi:10.1063/1.2001648 [48].
- Sidney Elmer, Sanghyun Park, & Vijay S. Pande (2005). "Foldamer dynamics expressed via Markov state models. II. State space decomposition". *Journal of Chemical Physics* **123**. doi:10.1063/1.2008230 [49].
- Sanghyun Park, Randall J. Radmer, Teri E. Klein, and Vijay S. Pande (2005). "A New Set of Molecular Mechanics Parameters for Hydroxyproline and Its Use in Molecular Dynamics Simulations of Collagen-Like Peptides". *Journal of Computational Chemistry* **26**: 1612–1616. doi:10.1002/jcc.20301 [50].
- Hideaki Fujutani, Yoshiaki Tanida, Masakatsu Ito, Guha Jayachandran, Christopher D. Snow, Michael R. Shirts, Eric J. Sorin, and Vijay S. Pande (2005). "Direct calculation of the binding free energies of FKBP ligands using the Fujitsu BioServer massively parallel computer". *Journal of Chemical Physics* **123**. doi:10.1063/1.1999637 [51].
- Nina Singhal and Vijay S. Pande (2005). "Error Analysis and efficient sampling in Markovian State Models for protein folding". *Journal of Chemical Physics* **123**. doi:10.1063/1.2116947 [52].
- Bojan Zagrovic, Guha Jayachandran, Ian S. Millett, Sebastian Doniach and Vijay S. Pande (2005). "How large is alpha-helix in solution? Studies of the radii of gyration of helical peptides by SAXS and MD". *Journal of Chemical Physics* **353**: 232–241. doi:10.1016/j.jmb.2005.08.053 [53].

2006

- Paula Petrone and Vijay S. Pande (2006). "Can conformational change be described by only a few normal modes?". *Biophysical Journal* **90**: 1583–1593. doi:10.1529/biophysj.105.070045 [54].
- Eric J. Sorin, Young Min Rhee, Michael R. Shirts, and Vijay S. Pande (2006). "The solvation interface is a determining factor in peptide conformational preferences". *Journal of Molecular Biology* **356**: 248–256. doi:10.1016/j.jmb.2005.11.058 [55].
- Eric J. Sorin and Vijay S. Pande (2006). "Nanotube confinement denatures protein helices". *Journal of the American Chemical Society* **128**: 6316–6317. doi:10.1021/ja060917j [56].
- Young Min Rhee and Vijay S. Pande (2006). "On the role of chemical detail in simulating protein folding kinetics". *Chemical Physics* **323**: 66–77. doi:10.1016/j.chemphys.2005.08.060 [57].
- L.T. Chong, W. C. Swope, J. W. Pitera, and V. S. Pande (2006). "A novel approach for computational alanine scanning: application to the p53 oligomerization domain". *Journal of Molecular Biology* **357** (3): 1039–1049. doi:10.1016/j.jmb.2005.12.083 [58].
- I. Suydam, C. D. Snow, V. S. Pande and S. G. Boxer. (2006). "Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory". *Science* **313** (5784): 200–204. doi:10.1126/science.1127159 [59].
- P. Kasson, N. Kelley, N. Singhal, M. Vrjlic, A. Brunger, and V. S. Pande (2006). "Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion". *Proceedings of the National Academy of Sciences, USA* **103** (32): 11916–11921. doi:10.1073/pnas.0601597103 [60].
- Guha Jayachandran, V. Vishal, and V. S. Pande (2006). "Folding Simulations of the Villin Headpiece in All-Atom Detail". *Journal of Chemical Physics* **124**. doi:10.1063/1.2186317 [61].

- Guha Jayachandran, M. R. Shirts, S. Park, and V. S. Pande (2006). "Parallelized Over Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics". *Journal of Chemical Physics* **125**. doi:10.1063/1.2221680 [62].
- C. Snow and V. S. Pande (2006). "Kinetic Definition of Protein Folding Transition State Ensembles and Reaction Coordinates". *Biophysical Journal* **91**: 14–24. doi:10.1529/biophysj.105.075689 [63].
- S. Park and V. S. Pande (2006). "A Bayesian Update Method for Adaptive Weighted Sampling". *Physical Review* **74** (6). doi:10.1103/PhysRevE.74.066703 [64].
- P. Kasson and V. S. Pande (2006). "Predicting structure and dynamics of loosely-ordered protein complexes: influenza hemagglutinin fusion peptide". *PSB*. doi:10.1142/9789812772435_0005 [65]. PMID 17992744.
- Erich Elsen, Mike Houston, V. Vishal, Eric Darve, Pat Hanrahan, and Vijay Pande (2006). "N-Body simulation on GPUs". *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. doi:10.1145/1188455.1188649 [66].

2007

- Guha Jayachandran, V. Vishal, Angel E. Garcıa and V. S. Pande (2007). "Local structure formation in simulations of two small proteins". *Journal of Structural Biology* **157** (3): 491–499. doi:10.1016/j.jsb.2006.10.001 [67].
- Adam L Beberg and Vijay S. Pande (2007). "Storage@home: Petascale Distributed Storage". *IPDPS*. doi:10.1109/IPDPS.2007.370672 [68].
- J. Chodera, N. Singhal, V. S. Pande, K. Dill, and W. Swope (2007). "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics". *Journal of Chemical Physics* **126** (15). PMID 17461665.
- D. Lucent, V. Vishal, V. S. Pande (2007). "Protein folding under confinement: a role for solvent". *Proceedings of the National Academy of Sciences, USA* **104** (25): 10430–10434. doi:10.1073/pnas.0608256104 [69].
- P. M. Kasson, A. Zomorodian, S. Park, N. Singhal, L. J. Guibas, and V. S. Pande (2007). "Persistent voids: a new structural metric for membrane fusion". *Bioinformatics*. doi:10.1093/bioinformatics/btm250 [70].
- P. M. Kasson and V. S. Pande (2007). "Control of Membrane Fusion Mechanism by Lipid Composition: Predictions from Ensemble Molecular Dynamics". *PLoS Computational Biology* **3** (11). doi:10.1371/journal.pcbi.0030220 [71].
- D. Ensign, P. M. Kasson, and V. S. Pande (2007). "Heterogeneity Even at the Speed Limit of Folding: Large-scale Molecular Dynamics Study of a Fast-folding Variant of the Villin Headpiece". *Journal of Molecular Biology* **374** (3): 806–816. doi:10.1016/j.jmb.2007.09.069 [72].
- Alex Robertson, Edgar Luttmann, Vijay S. Pande (2007). "Effects of long-range electrostatic forces on simulated protein folding kinetics". *Journal of Computational Chemistry* **29** (5): 694–700. doi:10.1002/jcc.20828 [73].
- Nina Singhal Hinrichs and Vijay S. Pande (2007). "Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics". *Journal of Chemical Physics* **126**. doi:10.1063/1.2740261 [74].

2008

- Xuhui Huang, Gregory R. Bowman,and Vijay S. Pande (2008). "Convergence of folding free energy landscapes via application of enhanced sampling methods in a distributed computing environment". *Journal of Chemical Physics* **128** (20). PMID 18513049.
- Gregory R. Bowman, Xuhui Huang, Yuan Yao, Jian Sun, Gunnar Carlsson, Leonidas J. Guibas, and Vijay S. Pande (2008). "Structural Insight into RNA Hairpin Folding Intermediates". *Journal of the American Chemical Society* **130** (30): 9676–9678. doi:10.1021/ja8032857 [75].
- Nicholas W. Kelley, V. Vishal, Grant A. Krafft, and Vijay S. Pande. (2008). "Simulating oligomerization at experimental concentrations and long timescales: A Markov state model approach.". *Journal of Chemical Physics* **129** (21). doi:10.1063/1.3010881 [76].
- Paula M. Petrone, Christopher D. Snow, Del Lucent, and Vijay S. Pande (2008). "Side-chain recognition and gating in the ribosome exit tunnel". *Proceedings of the National Academy of Sciences, USA* **105** (43): 16549–16554. doi:10.1073/pnas.0801795105 [77].
- Edgar Luttmann, Daniel L. Ensign, Vishal Vaidyanathan, Mike Houston, Noam Rimon, Jeppe Øland, Guha Jayachandran, Mark Friedrichs, Vijay S. Pande (2008). "Accelerating Molecular Dynamic Simulation on the Cell processor and PlayStation 3". *Journal of Computational Chemistry* **30** (2): 268–274. doi:10.1002/jcc.21054 [78].

2009

- Peter M. Kasson and Vijay S. Pande (2009). "Combining Mutual Information with Structural Analysis to Screen for Functionally Important Residues in Influenza Hemagglutinin". *Pacific Symposium on Biocomputing* **14**: 492–503. PMID 19209725.
- Nicholas W. Kelley, Xuhui Huang, Stephen Tam, Christoph Spiess, Judith Frydman and Vijay S. Pande (2009). "The predicted structure of the headpiece of the Huntingtin protein and its implications on Huntingtin aggregation". *Journal of Molecular Biology*. doi:10.1016/j.jmb.2009.01.032 [79].
- M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. LeGrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, V. S. Pande (2009). "Accelerating molecular dynamic simulation on graphics processing units". *Journal of Computational Chemistry*. doi:10.1002/jcc.21209 [80]. PMID 19191337.

High performance platforms

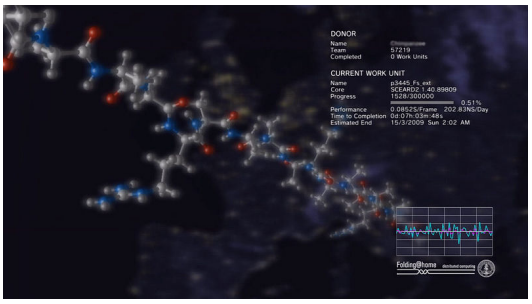
Graphical processing units

On October 2, 2006, the Folding@home Windows GPU client was released to the public as a beta test. After 9 days of processing from the Beta client the Folding@home project had received 31 teraFLOPs of computational performance from just 450 ATI Radeon X1900 GPUs, averaging at over 70x the performance of current CPU submissions, and the GPU clients remain the most powerful clients available in terms of performance per client (as of March 11, 2009, GPU clients accounted for over 60% of the entire project's throughput at an approximate ratio of 9 clients per teraFLOP—nVidia clients currently lead ATI clients in overall contribution and in performance per client).[5] On April 10, 2008, the second generation Windows GPU client was released to open beta testing, supporting ATI/AMD's Radeon HD 2000 and HD 3000 series, and also debuting a new core (GROGPU2 - Core 11).

Inaccuracies with DirectX were cited as the main reason for the migration to the new version (the original GPU client was officially retired June 6, 2008[81]), which uses AMD/ATI's CAL. On June 17, 2008, a version of the second-generation Windows GPU client for CUDA enabled Nvidia GPUs was also released for public beta testing.[82] The GPU clients proved reliable enough to be promoted out of the beta phase and were officially released August 1, 2008.[83] Newer GPU cores continue to be released for both CAL and CUDA. No word has to date been given over future support for OpenCL or DirectX 11's Compute Shaders.

While the only officially released GPU v2.0 client is for Windows, this client can be run on Linux under Wine with NVIDIA graphics cards.[84] The client can operate on both 32- and 64-bit Linux platforms, but in either case the 32-bit CUDA toolkit is required. This configuration is not officially supported, though initial results have shown comparable performance to that of the native client and no problems with the scientific results have been found . An unofficial installation guide has been published.[84]

PlayStation 3



The PlayStation 3's *Life With PlayStation* client replaced the Folding@home application on 18 September, 2008.

Stanford announced in August 2006 that a folding client was available to run on the Sony PlayStation 3.[85] The intent was that gamers would be able to contribute to the project by merely "contributing electricity", leaving their PlayStation 3 consoles running the client while not playing games. PS3 firmware version 1.6 (released on Thursday, March 22, 2007) allows for Folding@home software, a 50 MB download, to be used on the PS3.[5] A peak output of the project at 990 teraFLOPS was achieved on 25 March, 2007, at which time the number of FLOPS from each PS3 as reported by Stanford fell, reducing the overall speed rating of those machines by 50%. This had the effect of bumping down the overall project speed to the mid 700 range and increasing the number of active PS3s required to achieve a petaFLOPS level to around 60,000.

On April 26, 2007, Sony released a new version of Folding@home which improved folding performance drastically, such that the updated PS3 clients produced 1500 teraFLOPS with 52,000 clients versus the previous 400 teraFLOPS by around 24,000 clients.[86] Lately, the console accounts for around 26% of all teraFLOPS at an approximate ratio of 35½ PS3 clients per teraFLOPS.

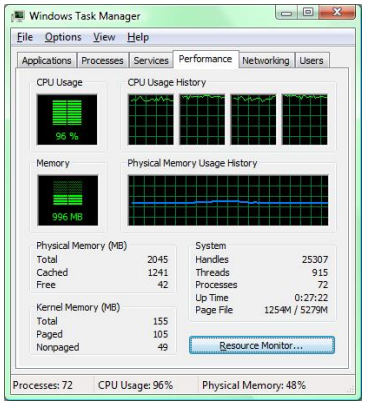
On December 19, 2007, Sony again updated the Folding@home client to version 1.3 to allow users to run music stored on their hard drives while contributing. Another feature of the 1.3 update allows users to automatically shut down their console after current work is done or after a limited period of time (for example 3 or 4 hours). Also, the software update added the Generalized Born implicit solvent model, so the FAH PS3 client gained more broad computing capabilities.[87] [88] Shortly afterward, 1.3.1 was released to solve a mishandling of protocol resulting in difficulties sending and receiving Work Units due to heavy server loads stemming from the fault.

On 18 September, 2008 the Folding@home client became *Life With PlayStation*. In addition to the existing functionality, the application also provides the user with access to information "channels", the first of which being the *Live Channel* which offers news headlines and weather through a 3D globe. The user can rotate and zoom in to any part of the world to access information provided by Google News and The Weather Channel, among other sources, all running whilst folding in the background. This update also provided more advanced simulation of protein folding and a new ranking system.[89]

Multi-core processing client

As more modern CPUs are being released, the migration to multiple cores is becoming more adopted by the public, and the Pande Group is adding symmetric multiprocessing (SMP) support to the Folding@home client in the hopes of capturing the additional processing power. The SMP support is being achieved by utilizing Message Passing Interface protocols. In current state it is being confined inside a single node by hard coded usage of the localhost.

On November 13, 2006, the beta SMP Folding@home clients for x86-64 Linux and x86 Mac OS X were released. The beta win32 SMP Folding@home client is out as well, and a 32-bit Linux client is currently in development.[90]



Folding@home SMP Client set to use 95% of a quad core processor.

Folding@home teams

A typical Folding@home user, running the client on a single PC, will likely not be ranked high on the list of contributors. However, if the user were to join a team, they would add the points they receive to a larger collective. Teams work by using the combined score of all their members. Thus, teams are ranked much higher than individual submitters. Rivalries between teams create friendly competition that benefits the folding community. Many teams publish their own stats, so members can have intra-team competitions for top spots.[91] Some teams offer prizes in an attempt to increase participation in the project.[92]

Development

The Folding@home project does not make the project source code available to the public, citing security and integrity concerns.[93] [94] At the same time, the majority of the scientific codes used by the FAH (ex. Cosm, GROMACS, TINKER, AMBER, CPMD, BrookGPU) are largely Open-source software or under similar licenses.

A development version of Folding@home once ran on the open source BOINC framework; however, this version remained unreleased.[95]

Estimated energy consumption

A PlayStation 3 has a maximum power rating of 380 watts. As Folding@home is a CPU intensive application, it causes 100% utilization. However, according to Stanford's PS3 FAQ, "We expect the PS3 to use about 200W while running Folding@home."[96] As of December 27, 2008, there are 55,291 PS3s providing 1,559,000,000 MFlops of processing power. This amounts to 28,196 MFlops/PS3, and with Stanford's estimate of 200W per PS3 (for original units manufactured on the 90nm process), 140.98 MFlops/watt.[5] This would put the PS3 portion of Folding@home at 95th on the November 2008 Green500 list.[97] The Cell processors used in current units of the PlayStation 3 utilize 65nm technology (lowering power consumption to around 115W per PS3), with another upgrade to 45nm planned (further dropping consumption to around 80W/PS3). This will further increase the power efficiency of the contribution from PlayStation 3 units.

The total power consumption required to produce the processing power required by the project can be estimated based upon the average FLOPS per watt. As of November 2008, according to the Green500 list, the most efficient computer - also based on a version of the Cell BE - runs at 536.24 MFLOPS/watt.[98] One petaFLOPS equals 1,000,000,000 MFLOPSs. Therefore, the current Folding@home project, if it were theoretically using the most efficient CPUs currently available, would use at least 2.8 megawatts of power per petaFLOPS, slightly more than the world's first and only petaflop system, the Cell-based Roadrunner which uses 2.345MW. This is equivalent to the power needed to light approximately 40,000 standard house light bulbs (between 60 and 100 watts each), or the equivalent of 0.5-3 electrical wind mills depending on their size.[99]

Estimates of power usage per time period are more difficult than estimates of power usage per processing instruction. This is because Folding@home clients are often run on computers that would be powered-on even in the absence of the Folding@home client, and that run other programs simultaneously. While Folding@home increases processor utilization, and thus (usually) power consumption, the extent to which it does so is dependent on the client processor's normal operating load, and its ability to reduce clock speeds when presented with less-than-full utilization (a process known as dynamic frequency scaling). Consequently, the total power usage of the Folding@home client on a temporal basis is probably less than the figure that could be calculated by summing the peak power consumption of each of the project's component processors.

See also

- Blue Gene
- Grid computing
- List of distributed computing projects
- Rosetta@Home
- Software for molecular modeling
- Molecular modeling on GPU

References

[1] http://www.scei.co.jp/folding/en/update.html
[2] http://folding.stanford.edu/English/License
[3] http://folding.stanford.edu
[4] Engadget, among other sites, announces that Guinness has recognized FAH as the most powerful distributed cluster, October 31, 2007. Retrieved November 5, 2007 (http://www.engadget.com/2007/10/31/folding-home-recognized-by-guinness-world-records/)
[5] http://fah-web.stanford.edu/cgi-bin/main.py?qtype=osstats|"Client Statistics by OS". *Folding@home distributed computing*. Stanford University. 2006-11-12 (updated automatically). http://fah-web.stanford.edu/cgi-bin/main.py?qtype=osstats. Retrieved on 2008-01-05.
[6] Vijay Pande (2006). http://folding.stanford.edu|"Folding@home distributed computing home page". Stanford University. http://folding.stanford.edu. Retrieved on 2006-11-12.
[7] http://folding.stanford.edu/FAQ-diseases.html|"Folding@home diseases studied FAQ". Stanford University. http://folding.stanford.edu/FAQ-diseases.html.
[8] http://twit.tv/fib27|"Futures in Biotech 27: Folding@home at 1.3 Petaflops" (Interview, webcast). http://twit.tv/fib27.
[9] http://folding.stanford.edu/English/About|"Folding@home - About" (FAQ). http://folding.stanford.edu/English/About.
[10] Vijay Pande and the Folding@home team (2009). http://folding.stanford.edu/English/Papers|"Folding@home - Papers". *Folding@home distributed computing*. Stanford University. http://folding.stanford.edu/English/Papers. Retrieved on 2009-02-19.
[11] C. Snow, H. Nguyen, V. S. Pande, and M. Gruebele. (2002). "Absolute comparison of simulated and experimental protein-folding dynamics". *Nature* **420** (6911): 102–106. doi: 10.1038/nature01160 (http://dx.doi.org/10.1038/nature01160). PMID 12422224.
[12] Vijay Pande (2005-10-16). http://folding.stanford.edu/QMD.html|"Folding@home with QMD core FAQ" (FAQ). Stanford University. http://folding.stanford.edu/QMD.html. Retrieved on 2006-12-03. *The site indicates that Folding@home uses a modification of CPMD allowing it to run on the supercluster environment.*
[13] http://fahwiki.net/index.php/Cores|"Cores - FaHWiki" (FAQ). http://fahwiki.net/index.php/Cores. Retrieved on 2007-11-06.
[14] http://foldingforum.org/viewtopic.php?f=52&t=8734&start=0|"Folding Forum: Announcing project 5900 and Core_14 on advmethods". 2009. http://foldingforum.org/viewtopic.php?f=52&t=8734&start=0. Retrieved on 2009-03-02.
[15] http://fahwiki.net/index.php/FAH_&_QMD_&_AMD64_&_SSE2|"FAH & QMD & AMD64 & SSE2" (FAQ). http://fahwiki.net/index.php/FAH_&_QMD_&_AMD64_&_SSE2.
[16] http://p450.caltech.edu/sharpen/sharpenabout.html|"SHARPEN: Systematic Hierarchical Algorithms for Rotamers and Proteins on an Extended Network" (About). http://p450.caltech.edu/sharpen/sharpenabout.html.
[17] http://p450.caltech.edu/sharpen/sharpenprojects.html|"SHARPEN". http://p450.caltech.edu/sharpen/sharpenprojects.html.
[18] Folding@home FLOP FAQ (http://folding.stanford.edu/English/FAQ-flops)
[19] http://forum.folding-community.org/fpost151025.html|"What is the state of Google Compute client?" (Blog). *Folding@home support forum*. Stanford University. http://forum.folding-community.org/fpost151025.html. Retrieved on 2006-11-12.
[20] Folding@home: Crossing the petaFLOPS barrier (http://folding.typepad.com/news/2007/09/crossing-the-pe.html)
[21] Folding@home: Post petaflop (http://folding.typepad.com/news/2007/09/post-petaflop.html)
[22] "Folding@home passes the 5 petaflop mark" (http://folding.typepad.com/news/2009/02/foldinghome-passes-the-5-petaflop-mark.html) from the official Folding@home blog
[23] http://www.top500.org/system/performance/9707|"TOP500 Roadrunner Performance Data". http://www.top500.org/system/performance/9707. Retrieved on 2008-12-27.
[24] Folding@home FLOP FAQ (http://folding.stanford.edu/English/FAQ-flops)
[25] http://dx.doi.org/10.1126%2Fscience.290.5498.1903
[26] http://dx.doi.org/10.1103%2FPhysRevLett.86.4983
[27] http://dx.doi.org/10.1006%2Fjmbi.2001.5033
[28] http://dx.doi.org/10.1016%2FS0022-2836%2802%2900997-X
[29] http://dx.doi.org/10.1016%2FS0022-2836%2802%2900888-4
[30] http://dx.doi.org/10.1021%2Fja028604l
[31] http://dx.doi.org/10.1002%2Fbip.10219

[32] http://dx.doi.org/10.1002%2Fjcc.10297
[33] http://dx.doi.org/10.1063%2F1.1587119
[34] http://dx.doi.org/10.1103%2FPhysRevLett.91.140601
[35] http://dx.doi.org/10.1038%2Fnsb995
[36] http://dx.doi.org/10.1016%2Fj.jmb.2004.02.024
[37] http://dx.doi.org/10.1073%2Fpnas.0305260101
[38] http://dx.doi.org/10.1073%2Fpnas.0307898101
[39] http://dx.doi.org/10.1063%2F1.1738647
[40] http://dx.doi.org/10.1016%2Fj.jmb.2004.10.083
[41] http://dx.doi.org/10.1529%2Fbiophysj.104.055087
[42] http://dx.doi.org/10.1529%2Fbiophysj.104.051938
[43] http://dx.doi.org/10.1002%2Fjcc.20208
[44] http://dx.doi.org/10.1146%2Fannurev.biophys.34.040204.144447
[45] http://dx.doi.org/10.1073%2Fpnas.0409693102
[46] http://dx.doi.org/10.1063%2F1.1873592
[47] http://dx.doi.org/10.1063%2F1.1877132
[48] http://dx.doi.org/10.1063%2F1.2001648
[49] http://dx.doi.org/10.1063%2F1.2008230
[50] http://dx.doi.org/10.1002%2Fjcc.20301
[51] http://dx.doi.org/10.1063%2F1.1999637
[52] http://dx.doi.org/10.1063%2F1.2116947
[53] http://dx.doi.org/10.1016%2Fj.jmb.2005.08.053
[54] http://dx.doi.org/10.1529%2Fbiophysj.105.070045
[55] http://dx.doi.org/10.1016%2Fj.jmb.2005.11.058
[56] http://dx.doi.org/10.1021%2Fja060917j
[57] http://dx.doi.org/10.1016%2Fj.chemphys.2005.08.060
[58] http://dx.doi.org/10.1016%2Fj.jmb.2005.12.083
[59] http://dx.doi.org/10.1126%2Fscience.1127159
[60] http://dx.doi.org/10.1073%2Fpnas.0601597103
[61] http://dx.doi.org/10.1063%2F1.2186317
[62] http://dx.doi.org/10.1063%2F1.2221680
[63] http://dx.doi.org/10.1529%2Fbiophysj.105.075689
[64] http://dx.doi.org/10.1103%2FPhysRevE.74.066703
[65] http://dx.doi.org/10.1142%2F9789812772435_0005
[66] http://dx.doi.org/10.1145%2F1188455.1188649
[67] http://dx.doi.org/10.1016%2Fj.jsb.2006.10.001
[68] http://dx.doi.org/10.1109%2FIPDPS.2007.370672
[69] http://dx.doi.org/10.1073%2Fpnas.0608256104
[70] http://dx.doi.org/10.1093%2Fbioinformatics%2Fbtm250
[71] http://dx.doi.org/10.1371%2Fjournal.pcbi.0030220
[72] http://dx.doi.org/10.1016%2Fj.jmb.2007.09.069
[73] http://dx.doi.org/10.1002%2Fjcc.20828
[74] http://dx.doi.org/10.1063%2F1.2740261
[75] http://dx.doi.org/10.1021%2Fja8032857
[76] http://dx.doi.org/10.1063%2F1.3010881
[77] http://dx.doi.org/10.1073%2Fpnas.0801795105
[78] http://dx.doi.org/10.1002%2Fjcc.21054
[79] http://dx.doi.org/10.1016%2Fj.jmb.2009.01.032
[80] http://dx.doi.org/10.1002%2Fjcc.21209
[81] http://folding.typepad.com/news/2008/06/gpu1-has-been-retired-gpu2-for-nvidia-release-nearing.html|"Folding@home: GPU1 has been retired, GPU2 for NVIDIA release nearing". http://folding.typepad.com/news/2008/06/gpu1-has-been-retired-gpu2-for-nvidia-release-nearing.html.
[82] http://folding.typepad.com/news/2008/06/gpu2-beta-client-for-nvidia-now-released.html|"Folding@home: GPU2 beta client for NVIDIA now released". http://folding.typepad.com/news/2008/06/gpu2-beta-client-for-nvidia-now-released.html.
[83] http://folding.typepad.com/news/2008/08/new-clients-are-out-620.html|"Folding@home: New clients are out (6.20)". http://folding.typepad.com/news/2008/08/new-clients-are-out-620.html.
[84] http://gpu2.twomurs.com/index.php?title=Main_Page|"Folding@Home GPU v2.0 Windows Client on Linux Wiki". 2008-08-23. http://gpu2.twomurs.com/index.php?title=Main_Page. Retrieved on 2008-11-06.

[85] Vijay Pande (2006-10-22). http://folding.stanford.edu/FAQ-PS3.html|"PS3 FAQ". Stanford University. http://folding.stanford.edu/FAQ-PS3.html. Retrieved on 2006-11-13.
[86] http://kotaku.com/gaming/folding@home/ps3-folding-kicking-ass-getting-update-255086.php|"PS3 Folding Kicking Ass, Getting Update". http://kotaku.com/gaming/folding%40home/ps3-folding-kicking-ass-getting-update-255086.php.
[87] http://www.scei.co.jp/folding/en/update.html|"Folding@home for PLAYSTATION3 Version 1.3". http://www.scei.co.jp/folding/en/update.html. Retrieved on 2007-12-31.
[88] Rimon, Noam (2007-12-18). http://blog.us.playstation.com/2007/12/18/new-foldinghome-features-coming/|"New Folding@home Features Coming". http://blog.us.playstation.com/2007/12/18/new-foldinghome-features-coming/. Retrieved on 2007-12-31.
[89] http://kotaku.com/5051551/life-with-playstation-out-now|"Life With PlayStation out now". Kotaku. http://kotaku.com/5051551/life-with-playstation-out-now.
[90] Vijay Pande (2006-11-13). http://folding.stanford.edu/FAQ-SMP.html|"Folding@home SMP Client FAQ". Stanford University. http://folding.stanford.edu/FAQ-SMP.html. Retrieved on 2006-11-13.
[91] [http://forum.folding-community.org/viewtopic.php?t=8846 Folding-community: why have teams?
[92] http://www.mprize.org/index.php?ctype=news&pagename=blogdetaildisplay&BID=2008032-20053630&detaildisplay=Y|"The Mprize-". http://www.mprize.org/index.php?ctype=news&pagename=blogdetaildisplay&BID=2008032-20053630&detaildisplay=Y.
[93] http://folding-community.org/viewtopic.php?p=178218&highlight=#178218|"Why not OpenSource?". http://folding-community.org/viewtopic.php?p=178218&highlight=#178218.
[94] http://folding.stanford.edu/English/FAQ-OpenSource|"Folding@home Open Source FAQ". http://folding.stanford.edu/English/FAQ-OpenSource.
[95] http://folding.stanford.edu/English/FAQ-highperformance|"FAH on BOINC". *Folding@home high performance client FAQ*. http://folding.stanford.edu/English/FAQ-highperformance.
[96] http://www.stanford.edu/group/pandegroup/folding/FAQ-PS3.html|"PS3 FAQ" (FAQ). http://www.stanford.edu/group/pandegroup/folding/FAQ-PS3.html.
[97] http://green500.org/lists/listdisplay.php?month=11&year=2008&list=green500_200811.csv&start=1&line=101|"Green 500". http://green500.org/lists/listdisplay.php?month=11&year=2008&list=green500_200811.csv&start=1&line=101.
[98] http://green500.org/lists/2008/11/list.php|"The Green500 List". http://green500.org/lists/2008/11/list.php. Retrieved on 2008-12-27.
[99] http://en.wikipedia.org/wiki/Windmill#Modern_Windmills|"Windmill - Wikipedia, the free encyclopedia". http://en.wikipedia.org/wiki/Windmill#Modern_Windmills.

External links

- Folding@home project homepage (http://folding.stanford.edu/)
- Folding@home Results (published papers) (http://folding.stanford.edu/Papers/)
- FAH blog (http://folding.typepad.com/)
- FAH Forum (http://foldingforum.org/)
- Folding@home Wiki (http://fahwiki.net/index.php/Main_Page)
- Official Folding@home Stats (http://folding.stanford.edu/English/Stats)
- Extreme OC Folding@home Stats (http://folding.ExtremeOverClocking.com/)
- Kakao Folding@home Stats (http://kakaostats.com/)
- Wikipedia team (http://fah-web.stanford.edu/cgi-bin/main.py?qtype=teampage&teamnum=42223)
- Massive folding farm (http://www.overclock.net/overclock-net-folding-home-team/370859-nitteo-s-f-h-gpu2-farm.html) Pics of a dedicated contributor's installation
- FoldWatcher (http://sourceforge.net/projects/foldwatcher) A Folding@Home monitoring application

Multi-media links

- Talk given by Folding@home author Vijay Pande at the PARC forum (http://www.parc.com/cms/get_article.php?id=799)
- Folding@home Instructional Video on YouTube (http://www.youtube.com/watch?v=2BVNCQt6MJw)
- Interview of Vijay Pande about Folding@Home Project (http://www.ustream.tv/recorded/1070617)

Molecular Dynamics, Theories and Computational Methods

Classical mechanics

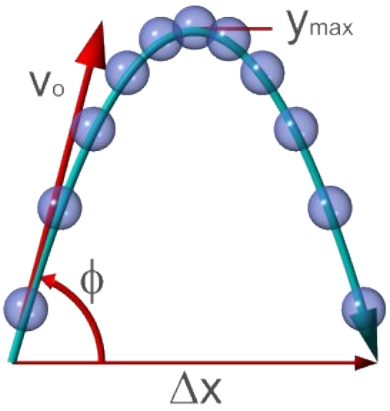
In physics, **classical mechanics** is one of the two major sub-fields of study in the science of mechanics, which is concerned with the set of physical laws governing and mathematically describing the motions of bodies and aggregates of bodies geometrically distributed within a certain boundary under the action of a system of forces. The other sub-field is quantum mechanics.

Classical mechanics is used for describing the motion of macroscopic objects, from projectiles to parts of machinery, as well as astronomical objects, such as spacecraft, planets, stars, and galaxies. It produces very accurate results within these domains, and is one of the oldest and largest subjects in science, engineering and technology.

Besides this, many related specialties exist, dealing with gases, liquids, and solids, and so on. Classical mechanics is enhanced by special relativity for objects moving with high velocity, approaching the speed of light; general relativity is employed to handle gravitation at a deeper level; and quantum mechanics handles the wave-particle duality of atoms and molecules.

The term *classical mechanics* was coined in the early 20th century to describe the system of mathematical physics begun by Isaac Newton and many contemporary 17th century natural philosophers, building upon the earlier astronomical theories of Johannes Kepler, which in turn were based on the precise observations of Tycho Brahe and the studies of terrestrial projectile motion of Galileo, but before the development of quantum physics and relativity. Therefore, some sources exclude so-called "relativistic physics" from that category. However, a number of modern sources *do* include Einstein's mechanics, which in their view represents *classical mechanics* in its most developed and most accurate form. The initial stage in the development of classical mechanics is often referred to as Newtonian mechanics, and is associated with the physical concepts employed by and the mathematical methods invented by Newton himself, in parallel with Leibniz, and others. This is further described in the following sections. More abstract and general methods include Lagrangian mechanics and Hamiltonian mechanics. Much of the content of classical mechanics was created in the 18th and 19th centuries and extends considerably beyond (particularly in its use of analytical mathematics) the work of Newton.

Description of the theory



The analysis of projectile motion is a part of classical mechanics.

The following introduces the basic concepts of classical mechanics. For simplicity, it often models real-world objects as point particles, objects with negligible size. The motion of a point particle is characterized by a small number of parameters: its position, mass, and the forces applied to it. Each of these parameters is discussed in turn.

In reality, the kind of objects which classical mechanics can describe always have a non-zero size. (The physics of *very* small particles, such as the electron, is more accurately described by quantum mechanics). Objects with non-zero size have more complicated behavior than hypothetical point particles, because of the additional degrees of freedom—for example, a baseball can spin while it is moving. However, the results for point particles can be used to study such objects by treating them as composite objects, made up of a large number of interacting point particles. The center of mass of a composite object behaves like a point particle.

Position and its derivatives

The SI derived "mechanical" (that is, not electromagnetic or thermal) units with kg, m and s	
Position	m
Angular position/Angle	unitless (radian)
velocity	$m\ s^{-1}$
Angular velocity	s^{-1}
acceleration	$m\ s^{-2}$
Angular acceleration	s^{-2}
jerk	$m\ s^{-3}$
"Angular jerk"	s^{-3}
specific energy	$m^2\ s^{-2}$
absorbed dose rate	$m^2\ s^{-3}$
moment of inertia	$kg\ m^2$
momentum	$kg\ m\ s^{-1}$
angular momentum	$kg\ m^2\ s^{-1}$
force	$kg\ m\ s^{-2}$
torque	$kg\ m^2\ s^{-2}$

energy	kg m^2 s^{-2}
power	kg m^2 s^{-3}
pressure and energy density	kg m^{-1} s^{-2}
surface tension	kg s^{-2}
Spring constant	kg s^{-2}
irradiance and energy flux	kg s^{-3}
kinematic viscosity	m^2 s^{-1}
dynamic viscosity	kg m^{-1} s
Density(mass density)	kg m^{-3}
Density(weight density)	kg m^{-2} s^{-2}
Number density	m^{-3}
Action	kg m^2 s^{-1}

The *position* of a point particle is defined with respect to an arbitrary fixed reference point, **O**, in space, usually accompanied by a coordinate system, with the reference point located at the *origin* of the coordinate system. It is defined as the vector **r** from **O** to the particle. In general, the point particle need not be stationary relative to **O**, so **r** is a function of *t*, the time elapsed since an arbitrary initial time. In pre-Einstein relativity (known as Galilean relativity), time is considered an absolute, i.e., the time interval between any given pair of events is the same for all observers. In addition to relying on absolute time, classical mechanics assumes Euclidean geometry for the structure of space.[1]

Velocity and speed

The *velocity*, or the rate of change of position with time, is defined as the derivative of the position with respect to time or

$$\vec{v} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}.$$

In classical mechanics, velocities are directly additive and subtractive. For example, if one car traveling East at 60 km/h passes another car traveling East at 50 km/h, then from the perspective of the slower car, the faster car is traveling east at 60 − 50 = 10 km/h. Whereas, from the perspective of the faster car, the slower car is moving 10 km/h to the West. Velocities are directly additive as vector quantities; they must be dealt with using vector analysis.

Mathematically, if the velocity of the first object in the previous discussion is denoted by the vector $\vec{u} = u\vec{d}$ and the velocity of the second object by the vector $\vec{v} = v\vec{e}$ where u is the speed of the first object, v is the speed of the second object, and $\vec{d}$ and $\vec{e}$ are unit vectors in the directions of motion of each particle respectively, then the velocity of the first object as seen by the second object is:

$$\vec{u'} = \vec{u} - \vec{v}$$

Similarly:

$$\vec{v'} = \vec{v} - \vec{u}$$

When both objects are moving in the same direction, this equation can be simplified to:

$$\vec{u'} = (u - v)\vec{d}$$

Or, by ignoring direction, the difference can be given in terms of speed only:

$$u' = u - v$$

Acceleration

The *acceleration*, or rate of change of velocity, is the derivative of the velocity with respect to time (the second derivative of the position with respect to time) or

$$\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t}.$$

Acceleration can arise from a change with time of the magnitude of the velocity or of the direction of the velocity or both. If only the magnitude, v, of the velocity decreases, this is sometimes referred to as *deceleration*, but generally any change in the velocity with time, including deceleration, is simply referred to as acceleration.

Frames of reference

While the position and velocity and acceleration of a particle can be referred to any observer in any state of motion, classical mechanics assumes the existence of a special family of reference frames in terms of which the mechanical laws of nature take a comparatively simple form. These special reference frames are called inertial frames. They are characterized by the absence of acceleration of the observer and the requirement that all forces entering the observer's physical laws originate in identifiable sources (charges, gravitational bodies, and so forth). A non-inertial reference frame is one accelerating with respect to an inertial one, and in such a non-inertial frame a particle is subject to acceleration by fictitious forces that enter the equations of motion solely as a result of its accelerated motion, and do not originate in identifiable sources. These fictitious forces are in addition to the real forces recognized in an inertial frame. A key concept of inertial frames is the method for identifying them. (See inertial frame of reference for a discussion.) For practical purposes, reference frames that are unaccelerated with respect to the distant stars are regarded as good approximations to inertial frames.

The following consequences can be derived about the perspective of an event in two inertial reference frames, S and S', where S' is traveling at a relative velocity of $\vec{u}$ to S.

- $\vec{v'}=\vec{v}-\vec{u}$ (the velocity $\vec{v'}$ of a particle from the perspective of *S*' is slower by $\vec{u}$ than its velocity $\vec{v}$ from the perspective of *S*)
- $\vec{a'}=\vec{a}$ (the acceleration of a particle is the same in any inertial reference frame)
- $\vec{F'}=\vec{F}$ (the force on a particle is the same in any inertial reference frame)
- the speed of light is not a constant in classical mechanics, nor does the special position given to the speed of light in relativistic mechanics have a counterpart in classical mechanics.
- the form of Maxwell's equations is not preserved across such inertial reference frames. However, in Einstein's theory of special relativity, the assumed constancy (invariance) of the vacuum speed of light alters the relationships between inertial reference frames so as to render Maxwell's equations invariant.

Forces; Newton's Second Law

Newton was the first to mathematically express the relationship between force and momentum. Some physicists interpret Newton's second law of motion as a definition of force and mass, while others consider it to be a fundamental postulate, a law of nature. Either interpretation has the same mathematical consequences, historically known as "Newton's Second Law":

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \frac{\mathrm{d}(m\vec{v})}{\mathrm{d}t}.$$

The quantity $m\vec{v}$ is called the (canonical) momentum. The net force on a particle is thus equal to rate change of momentum of the particle with time. Since the definition of acceleration is $\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t}$, the second law can be written in the simplified and more familiar form

$$\vec{F} = m\vec{a}\,.$$

So long as the force acting on a particle is known, Newton's second law is sufficient to describe the motion of a particle. Once independent relations for each force acting on a particle are available, they can be substituted into Newton's second law to obtain an ordinary differential equation, which is called the *equation of motion*.

As an example, assume that friction is the only force acting on the particle, and that it may be modeled as a function of the velocity of the particle, for example:

$$\vec{F}_{\mathrm{R}} = -\lambda\vec{v}$$

with λ a positive constant. Then the equation of motion is

$$-\lambda\vec{v} = m\vec{a} = m\frac{\mathrm{d}\vec{v}}{\mathrm{d}t}.$$

This can be integrated to obtain

$$\vec{v} = \vec{v}_0 e^{-\lambda t/m}$$

where $\vec{v}_0$ is the initial velocity. This means that the velocity of this particle decays exponentially to zero as time progresses. In this case, an equivalent viewpoint is that the kinetic energy of the particle is absorbed by friction (which converts it to heat energy in accordance with the conservation of energy), slowing it down. This expression can be further integrated to obtain the position $\vec{r}$ of the particle as a function of time.

Important forces include the gravitational force and the Lorentz force for electromagnetism. In addition, Newton's third law can sometimes be used to deduce the forces acting on a particle: if it is known that particle A exerts a force $\vec{F}$ on another particle B, it follows that B must exert an equal and opposite *reaction force*, $-\vec{F}$, on A. The strong form of Newton's third law requires that $\vec{F}$ and $-\vec{F}$ act along the line connecting A and B, while the weak form does not. Illustrations of the weak form of Newton's third law are often found for magnetic forces.

Energy

If a force $\vec{F}$ is applied to a particle that achieves a displacement $\Delta\vec{r}$, the *work done* by the force is defined as the scalar product of force and displacement vectors: (noting that the displacement vector is the change in position vector)

$$W = \vec{F} \cdot \Delta\vec{r}.$$

If the mass of the particle is constant, and W_{total} is the total work done on the particle, obtained by summing the work done by each applied force, from Newton's second law:

$$W_{\text{total}} = \Delta E_k,$$

where E_k is called the kinetic energy. For a point particle, it is mathematically defined as the amount of work done to accelerate the particle from zero velocity to the given velocity v:

$$E_k = \tfrac{1}{2}mv^2.$$

For extended objects composed of many particles, the kinetic energy of the composite body is the sum of the kinetic energies of the particles.

A particular class of forces, known as *conservative forces*, can be expressed as the gradient of a scalar function, known as the potential energy and denoted E_p:

$$\vec{F} = -\vec{\nabla} E_p.$$

If all the forces acting on a particle are conservative, and E_p is the total potential energy (which is defined as a work of involved forces to rearrange mutual positions of bodies), obtained by summing the potential energies corresponding to each force

$$\vec{F} \cdot \Delta\vec{r} = -\vec{\nabla} E_p \cdot \Delta\vec{s} = -\Delta E_p \Rightarrow -\Delta E_p = \Delta E_k \Rightarrow \Delta(E_k + E_p) = 0.$$

This result is known as *conservation of energy* and states that the total energy,

$$\sum E = E_k + E_p$$

is constant in time. It is often useful, because many commonly encountered forces are conservative.

Beyond Newton's Laws

Classical mechanics also includes descriptions of the complex motions of extended non-pointlike objects. Euler's laws provide extensions to Newton's laws in this area. The concepts of angular momentum rely on the same calculus used to describe one-dimensional motion.

There are two important alternative formulations of classical mechanics: Lagrangian mechanics and Hamiltonian mechanics. These, and other modern formulations, usually bypass the concept of "force", instead referring to other physical quantities, such as energy, for describing mechanical systems.

Classical transformations

Consider two reference frames S and S' . For observers in each of the reference frames an event has space-time coordinates of (x,y,z,t) in frame S and (x' ,y' ,z' ,t') in frame S' . Assuming time is measured the same in all reference frames, and if we require $x = x'$ when $t = 0$, then the relation between the space-time coordinates of the same event observed from the reference frames S' and S, which are moving at a relative velocity of u in the x

direction is:

$x' = x - ut$

$y' = y$

$z' = z$

$t' = t$

This set of formulas defines a group transformation known as the Galilean transformation (informally, the *Galilean transform*). This group is a limiting case of the Poincaré group used in special relativity. The limiting case applies when the velocity u is very small compared to c, the speed of light.

For some problems, it is convenient to use rotating coordinates (reference frames). Thereby one can either keep a mapping to a convenient inertial frame, or introduce additionally a fictitious centrifugal force and Coriolis force.

History

Some Greek philosophers of antiquity, among them Aristotle, may have been the first to maintain the idea that "everything happens for a reason" and that theoretical principles can assist in the understanding of nature. While to a modern reader, many of these preserved ideas come forth as eminently reasonable, there is a conspicuous lack of both mathematical theory and controlled experiment, as we know it. These both turned out to be decisive factors in forming modern science, and they started out with classical mechanics.

An early experimental scientific method was introduced into mechanics in the 11th century by al-Biruni, who along with al-Khazini in the 12th century, unified statics and dynamics into the science of mechanics, and combined the fields of hydrostatics with dynamics to create the field of hydrodynamics.[2] Concepts related to Newton's laws of motion were also enunciated by several other Muslim physicists during the Middle Ages. Early versions of the law of inertia, known as Newton's first law of motion, and the concept relating to momentum, part of Newton's second law of motion, were described by Ibn al-Haytham (Alhacen)[3] [4] and Avicenna.[5] [6] The proportionality between force and acceleration, an important principle in classical mechanics, was first stated by Hibat Allah Abu'l-Barakat al-Baghdaadi,[7] and theories on gravity were developed by Ja'far Muhammad ibn Mūsā ibn Shākir,[8] Ibn al-Haytham,[9] and al-Khazini.[10] It is known that Galileo Galilei's mathematical treatment of acceleration and his concept of impetus[11] grew out of earlier medieval analyses of motion, especially those of Avicenna,[5] Ibn Bajjah,[12] and Jean Buridan.

The first published causal explanation of the motions of planets was Johannes Kepler's Astronomia nova published in 1609. He concluded, based on Tycho Brahe's observations of the orbit of Mars, that the orbits were ellipses. This break with ancient thought was happening around the same time that Galilei was proposing abstract mathematical laws for the motion of objects. He may (or may not) have performed the famous experiment of dropping two cannon balls of different masses from the tower of Pisa, showing that they both hit the ground at the same time. The reality of this experiment is disputed, but, more importantly, he did carry out quantitative experiments by rolling balls on an inclined plane. His theory of accelerated motion derived from the results of such experiments, and forms a cornerstone of classical mechanics.

As foundation for his principles of natural philosophy, Newton proposed three laws of motion: the law of inertia, his second law of acceleration (mentioned above), and the law of action and reaction; and hence laid the foundations for classical mechanics. Both Newton's second and third laws were given proper scientific and mathematical treatment in Newton's Philosophiæ Naturalis Principia Mathematica, which distinguishes them from earlier attempts at explaining similar phenomena, which were either incomplete, incorrect, or given little accurate mathematical expression. Newton also enunciated the principles of conservation of momentum and angular momentum. In Mechanics, Newton was also the first to provide the first correct scientific and mathematical formulation of gravity in Newton's law of universal gravitation. The combination of Newton's laws of motion and gravitation provide the fullest and most accurate description of classical mechanics. He demonstrated that these laws apply to everyday objects as well as to celestial objects. In particular, he obtained a theoretical explanation of Kepler's laws of motion of the planets.

Newton previously invented the calculus, of mathematics, and used it to perform the mathematical calculations. For acceptability, his book, the Principia, was formulated entirely in terms of the long established geometric methods, which were soon to be eclipsed by his calculus. However it was Leibniz who developed the notation of the derivative and integral preferred today.

Newton, and most of his contemporaries, with the notable exception of Huygens, worked on the assumption that classical mechanics would be able to explain all phenomena, including light, in the form of geometric optics. Even when discovering the so-called Newton's rings (a wave interference phenomenon) his explanation remained with his own corpuscular theory of light.

After Newton, classical mechanics became a principal field of study in mathematics as well as physics.

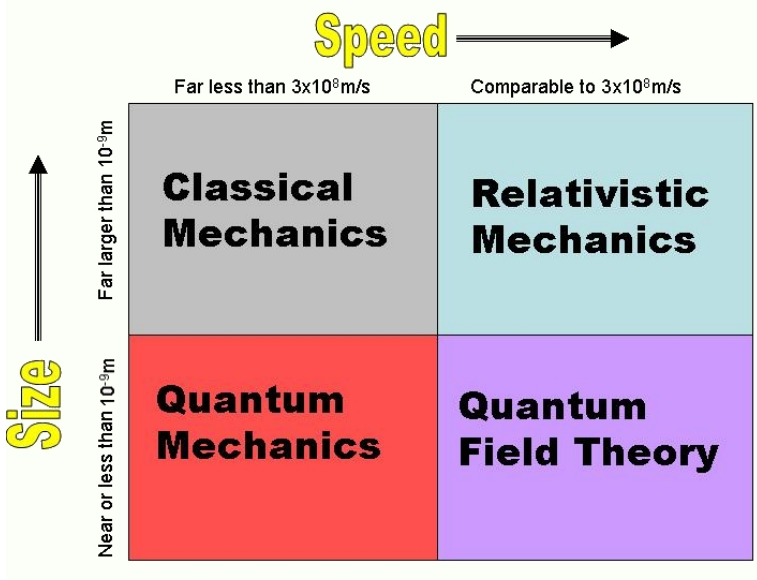
Some difficulties were discovered in the late 19th century that could only be resolved by more modern physics. Some of these difficulties related to compatibility with electromagnetic theory, and the famous Michelson-Morley experiment. The resolution of these problems led to the special theory of relativity, often included in the term classical mechanics.

A second set of difficulties were related to thermodynamics. When combined with thermodynamics, classical mechanics leads to the Gibbs paradox of classical statistical mechanics, in which entropy is not a well-defined quantity. Black-body radiation was not explained without the introduction of quanta. As experiments reached the atomic level, classical mechanics failed to explain, even approximately, such basic things as the energy levels and sizes of atoms and the photo-electric effect. The effort at resolving these problems led to the development of quantum mechanics.

Since the end of the 20th century, the place of classical mechanics in physics has been no longer that of an independent theory. Emphasis has shifted to understanding the fundamental forces of nature as in the Standard model and its more modern extensions into a unified theory of everything.[13] Classical mechanics is a theory for the study of the motion of non-quantum mechanical, low-energy particles in weak gravitational fields.

Limits of validity

Many branches of classical mechanics are simplifications or approximations of more accurate forms; two of the most accurate being general relativity and relativistic statistical mechanics. Geometric optics is an approximation to the quantum theory of light, and does not have a superior "classical" form.



Domain of validity for Classical Mechanics

The Newtonian approximation to special relativity

Newtonian, or non-relativistic classical momentum

$$\vec{p} = m_0\vec{v}$$

is the result of the first order Taylor approximation of the relativistic expression:

$$\vec{p} = \frac{m_0\vec{v}}{\sqrt{1 - v^2/c^2}} = m_0\vec{v}\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \cdots\right), \text{ where } v = |\vec{v}|$$

when expanded about

$$\frac{v}{c} = 0$$

so it is only valid when the velocity is much less than the speed of light. Quantitatively speaking, the approximation is good so long as

$$\left(\frac{v}{c}\right)^2 << 1$$

For example, the relativistic cyclotron frequency of a cyclotron, gyrotron, or high voltage magnetron is given by $f = f_c\frac{m_0}{m_0 + T/c^2}$, where $f_c$is the classical frequency of an electron (or other charged particle) with kinetic energy $T$and (rest) mass $m_0$circling in a magnetic field. The (rest) mass of an electron is 511 keV. So the frequency correction is 1% for a magnetic vacuum tube with a 5.11 kV. direct current accelerating voltage.

The classical approximation to quantum mechanics

The ray approximation of classical mechanics breaks down when the de Broglie wavelength is not much smaller than other dimensions of the system. For non-relativistic particles, this wavelength is

$$\lambda = \frac{h}{p}$$

where *h* is Planck's constant and *p* is the momentum.

Again, this happens with electrons before it happens with heavier particles. For example, the electrons used by Clinton Davisson and Lester Germer in 1927, accelerated by 54 volts, had a wave length of 0.167 nm, which was long enough to exhibit a single diffraction side lobe when reflecting from the face of a nickel crystal with atomic spacing of 0.215 nm. With a larger vacuum chamber, it would seem relatively easy to increase the angular resolution from around a radian to a milliradian and see quantum diffraction from the periodic patterns of integrated circuit computer memory.

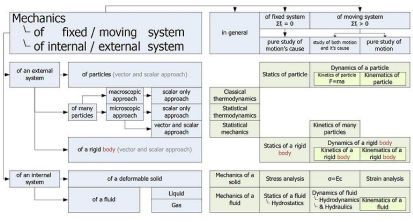
More practical examples of the failure of classical mechanics on an engineering scale are conduction by quantum tunneling in tunnel diodes and very narrow transistor gates in integrated circuits.

Classical mechanics is the same extreme high frequency approximation as geometric optics. It is more often accurate because it describes particles and bodies with rest mass. These have more momentum and therefore shorter De Broglie wavelengths than massless particles, such as light, with the same kinetic energies.

Branches

Classical mechanics was traditionally divided into three main branches:

- Statics, the study of equilibrium and its relation to forces
- Dynamics, the study of motion and its relation to forces
- Kinematics, dealing with the implications of observed motions without regard for circumstances causing them



Branches of mechanics

Another division is based on the choice of mathematical formalism:

- Newtonian mechanics
- Lagrangian mechanics
- Hamiltonian mechanics

Alternatively, a division can be made by region of application:

- Celestial mechanics, relating to stars, planets and other celestial bodies
- Continuum mechanics, for materials which are modelled as a continuum, e.g., solids and fluids (i.e., liquids and gases).
- Relativistic mechanics (i.e. including the special and general theories of relativity), for bodies whose speed is close to the speed of light.

- Statistical mechanics, which provides a framework for relating the microscopic properties of individual atoms and molecules to the macroscopic or bulk thermodynamic properties of materials.

See also

- History of classical mechanics
- Dynamical systems
- List of equations in classical mechanics
- List of publications in classical mechanics
- Molecular dynamics
- Newton's laws of motion
- Special theory of relativity

Notes

[1] MIT physics 8.01 lecture notes (page 12) (http://ocw.mit.edu/NR/rdonlyres/Physics/8-01Physics-IFall2003/B4144452-A6DE-464D-A0FA-D4D057AA9222/0/binder1.pdf) (PDF)

[2] Mariam Rozhanskaya and I. S. Levinova (1996), "Statics", in Roshdi Rashed, ed., *Encyclopedia of the History of Arabic Science*, Vol. 2, p. 614-642 [642], Routledge, London and New York

[3] Abdus Salam (1984), "Islam and Science". In C. H. Lai (1987), *Ideals and Realities: Selected Essays of Abdus Salam*, 2nd ed., World Scientific, Singapore, p. 179-213.

[4] Seyyed Hossein Nasr, "The achievements of Ibn Sina in the field of science and his contributions to its philosophy", *Islam & Science*, December 2003.

[5] Fernando Espinoza (2005). "An analysis of the historical development of ideas about motion and its implications for teaching", *Physics Education* **40** (2), p. 141.

[6] Seyyed Hossein Nasr, "Islamic Conception Of Intellectual Life", in Philip P. Wiener (ed.), *Dictionary of the History of Ideas*, Vol. 2, p. 65, Charles Scribner's Sons, New York, 1973-1974.

[7] Shlomo Pines (1970). "Abu'l-Barakāt al-Baghdādī, Hibat Allah". *Dictionary of Scientific Biography*. **1**. New York: Charles Scribner's Sons. pp. 26-28. ISBN 0684101149.
(cf. Abel B. Franco (October 2003). "Avempace, Projectile Motion, and Impetus Theory", *Journal of the History of Ideas* **64** (4), p. 521-546 [528]

[8] Robert Briffault (1938). *The Making of Humanity*, p. 191.

[9] Nader El-Bizri (2006), "Ibn al-Haytham or Alhazen", in Josef W. Meri (2006), *Medieval Islamic Civilization: An Encyclopaedia*, Vol. II, p. 343-345, Routledge, New York, London.

[10] Mariam Rozhanskaya and I. S. Levinova (1996), "Statics", in Roshdi Rashed, ed., *Encyclopaedia of the History of Arabic Science*, Vol. 2, p. 622. London and New York: Routledge.

[11] Galileo Galilei, *Two New Sciences*, trans. Stillman Drake, (Madison: Univ. of Wisconsin Pr., 1974), pp 217, 225, 296-7.

[12] Ernest A. Moody (1951). "Galileo and Avempace: The Dynamics of the Leaning Tower Experiment (I)", *Journal of the History of Ideas* **12** (2), p. 163-193.

[13] Page 2-10 of the *Feynman Lectures on Physics* says "For already in classical mechanics there was indeterminability from a practical point of view." The past tense here implies that classical physics is no longer fundamental.

References

- Feynman, Richard (1996). *Six Easy Pieces*. Perseus Publishing. ISBN 0-201-40825-2.
- Feynman, Richard; Phillips, Richard (1998). *Six Easy Pieces*. Perseus Publishing. ISBN 0-201-32841-0.
- Feynman, Richard (1999). *Lectures on Physics*. Perseus Publishing. ISBN 0-7382-0092-1.
- Landau, L. D.; Lifshitz, E. M. (1972). *Mechanics Course of Theoretical Physics , Vol. 1*. Franklin Book Company. ISBN 0-08-016739-X.
- Kleppner, D. and Kolenkow, R. J., *An Introduction to Mechanics*, McGraw-Hill (1973). ISBN 0-07-035048-5
- Gerald Jay Sussman and Jack Wisdom, *Structure and Interpretation of Classical Mechanics*, MIT Press (2001). ISBN 0-262-19455-4}
- Herbert Goldstein, Charles P. Poole, John L. Safko, *Classical Mechanics (3rd Edition)*, Addison Wesley; ISBN 0-201-65702-3
- Robert Martin Eisberg, *Fundamentals of Modern Physics*, John Wiley and Sons, 1961
- M. Alonso, J. Finn, "Fundamental university physics", Addison-Wesley

External links

- Crowell, Benjamin. Newtonian Physics (http://www.lightandmatter.com/area1book1.html) (an introductory text, uses algebra with optional sections involving calculus)
- Fitzpatrick, Richard. Classical Mechanics (http://farside.ph.utexas.edu/teaching/301/301.html) (uses calculus)
- Hoiland, Paul (2004). Preferred Frames of Reference & Relativity (http://doc.cern.ch//archive/electronic/other/ext/ext-2004-126.pdf)
- Horbatsch, Marko, " *Classical Mechanics Course Notes* (http://www.yorku.ca/marko/PHYS2010/index.htm)".
- Rosu, Haret C., " *Classical Mechanics* (http://arxiv.org/abs/physics/9909035)". Physics Education. 1999. [arxiv.org : physics/9909035]
- Schiller, Christoph. Motion Mountain (http://www.motionmountain.net) (an introductory text, uses some calculus; see also Motion Mountain)
- Sussman, Gerald Jay & Wisdom, Jack & Mayer,Meinhard E. (2001). Structure and Interpretation of Classical Mechanics (http://mitpress.mit.edu/SICM/)
- Tong, David. Classical Dynamics (http://www.damtp.cam.ac.uk/user/tong/dynamics.html) (Cambridge lecture notes on Lagrangian and Hamiltonian formalism)
- Kinematic Models for Design Digital Library (KMODDL) (http://kmoddl.library.cornell.edu/index.php)
  Movies and photos of hundreds of working mechanical-systems models at Cornell University. Also includes an e-book library (http://kmoddl.library.cornell.edu/e-books.php) of classic texts on mechanical design and engineering.

Newton's laws of motion

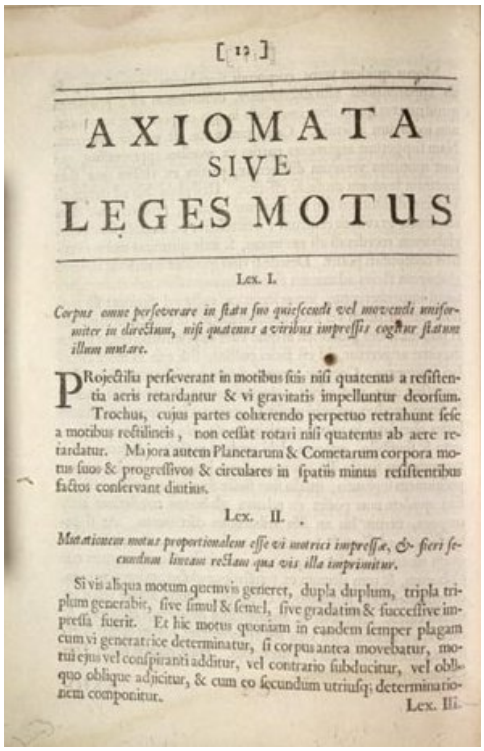
Newton's laws of motion are three physical laws that form the basis for classical mechanics. They are:

1. A body at rest stays at rest, and a body in motion stays in motion, unless it is acted on by an external force.

2. Force equals mass times acceleration (**F** = m**a**) (or alternately, force equals the time rate of change of momentum).

3. To every action there is an equal and opposite reaction.

They describe the relationship between the forces acting on a body to the motion of the body. They were first compiled by Sir Isaac Newton in his work *Philosophiæ Naturalis Principia Mathematica*, first published on July 5, 1687.[1] Newton used them to explain and investigate the motion of many physical objects and systems.[2] For example, in the third volume of the text, Newton showed that these laws of motion, combined with his law of universal gravitation, explained Kepler's laws of planetary motion.



[13]

AXIOMATA
SIVE
LEGES MOTUS

Lex. I.

Corpus omne perſeverare in ſtatu ſuo quieſcendi vel movendi uniformiter in directum, niſi quatenus a viribus impreſſis cogitur ſtatum illum mutare.

Projectilia perſeverant in motibus ſuis niſi quatenus a reſiſtentia aeris retardantur & vi gravitatis impelluntur deorſum. Trochus, cujus partes cohærendo perpetuo retrahunt ſeſe a motibus rectilineis, non ceſſat rotari niſi quatenus ab aere retardatur. Majora autem Planetarum & Cometarum corpora motus ſuos & progreſſivos & circulares in ſpatiis minus reſiſtentibus factos conſervant diutius.

Lex. II.

Mutationem motus proportionalem eſſe vi motrici impreſſæ, & fieri ſecundum lineam rectam qua vis illa imprimitur.

Si vis aliqua motum quemvis generet, dupla duplum, tripla triplum generabit, ſive ſimul & ſemel, ſive gradatim & ſucceſſive impreſſa fuerit. Et hic motus quoniam in eandem ſemper plagam cum vi generatrice determinatur, ſi corpus antea movebatur, motui ejus vel conſpiranti additur, vel contrario ſubducitur, vel obliquo obliquè adjicitur, & cum eo ſecundum utriuſq; determinationem componitur.

Lex. III.

Newton's First and Second laws, in Latin, from the original 1687 edition of the Principia Mathematica.

The three laws

First law

There exists a set of inertial reference frames relative to which all particles with no net force acting on them will move without change in their velocity. This law is often simplified as "A body persists its state of rest or of uniform motion unless acted upon by an external unbalanced force." Newton's first law is often referred to as the law of inertia.

Second law

Observed from an inertial reference frame, the net force on a particle is proportional to the time rate of change of its linear momentum: $\mathbf{F} = \mathrm{d}(m\mathbf{v})/\mathrm{d}t$. This law is often stated as, "Force equals mass times acceleration ($\mathbf{F} = m\mathbf{a}$)": the net force on an object is equal to the mass of the object multiplied by its acceleration.

Third law

Whenever a particle A exerts a force on another particle B, B simultaneously exerts a force on A with the same magnitude in the opposite direction. The strong form of the law further postulates that these two forces act along the same line. This law is often simplified into the sentence, "To every action there is an equal and opposite reaction."

In the given interpretation mass, acceleration, momentum, and (most importantly) force are assumed to be externally defined quantities. This is the most common, but not the only interpretation: one can consider the laws to be a definition of these quantities. Notice that the second law only holds when the observation is made from an inertial reference frame, and since an inertial reference frame is defined by the first law, asking a proof of the first law from the second law is a logical fallacy. At speeds approaching the speed of light the effects of special relativity must be taken into account.[3]

Newton's first law: law of inertia

> Lex I: *Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare.* **Every body persists in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by force impressed.**[4]

Newton's first law is also called the **law of inertia**. In a simplified form, it states that if the vector sum of all forces (also known as the net force) acting on an object is zero, then the state of motion of the object does not change. In particular: Newton's first law: An object at rest remains at rest and an object in motion will remain in motion unless acted on by an unbalanced force.

- An object that is not moving will not move until a net force acts upon it.
- An object that is moving will not change its velocity (accelerate) until a net force acts upon it.

The first point needs no comment, but the second seems to violate everyday experience. A hockey puck sliding along a table doesn't move forever; rather, it slows and eventually comes to a stop. According to Newton's laws, though, the hockey puck does not stop of its own accord, but because of a force applied in the opposite direction to the direction of motion. That force is easily identified as a frictional force between the table and the puck. In the absence of such a force, as approximated by an air hockey table or ice rink, the puck's motion would not slow.

There are no perfect demonstrations of the law, as friction usually causes a force to act on a moving body, and even in outer space gravitational forces act and cannot be shielded against, but the law serves to emphasize the elementary causes of changes in an object's state of motion.

The above treatment of Newton's first law is an over-simplification, though. A more sophisticated approach to the law of inertia is given by:

> **There is a class of frames of reference (called *inertial frames*) relative to which the motion of a particle not subject to forces is a straight line.**[5]

Newton placed the law of inertia first to establish frames of reference for which the other laws are applicable (see Gailili & Tseitlin,[6] or Woodhouse[5]). Such frames are called inertial frames.

To understand why the laws are restricted to inertial frames, consider a ball at rest within an accelerating body: an airplane on a runway will suffice for this example. From the perspective of anyone within the airplane (that is, from the airplane's *frame of reference* when put in technical terms) the ball will appear to move backwards as the plane accelerates forwards (the same feeling as being pushed back into your seat as the plane

accelerates). This motion appears to contradict Newton's second law as, from the point of view of the passengers, there appears to be no force acting on the ball that would cause it to move. The reason why there is in fact no contradiction to the second law is because Newton's *second* law (without modification) is not applicable in this situation: Newton's *first* law does not apply because the stationary ball *does not* remain stationary. Thus, it is important to establish whether the various laws are applicable or not, inasmuch as they are not applicable in all situations.[7]

History of the Law of Inertia

Newton's first law is a restatement of what Galileo had already described and Newton gave credit to Galileo. It differs from Aristotle's view that all objects have a natural place in the universe. Aristotle believed that heavy objects like rocks wanted to be at rest on the Earth and that light objects like smoke wanted to be at rest in the sky and the stars wanted to remain in the heavens. However, a key difference between Galileo's idea and Aristotle's is that Galileo realized that force acting on a body determines *acceleration*, not velocity. This insight leads to Newton's First Law—no force means no acceleration, and hence the body will maintain its velocity.

The law of inertia apparently occurred to several different natural philosophers and scientists independently. The inertia of motion was described in the 3rd century BC by the Chinese philosopher Mo Tzu, and in the 11th century by the Muslim scientists, Alhazen[8] and Avicenna.[9] The 17th century philosopher René Descartes also formulated the law, although he did not perform any experiments to confirm it.

Newton's second law

> Lex II: *Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.*
>
> **The change of momentum of a body is proportional to the impulse impressed on the body, and happens along the straight line on which that impulse is impressed.**

In Motte's 1729 translation (from Newton's Latin), the second law of motion reads:[10]

> LAW II: The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed. — If a force generates a motion, a double force will generate double the motion, a triple force triple the motion, whether that force be impressed altogether and at once, or gradually and successively. And this motion (being always directed the same way with the generating force), if the body moved before, is added to or subtracted from the former motion, according as they directly conspire with or are directly contrary to each other; or obliquely joined, when they are oblique, so as to produce a new motion compounded from the determination of both.

Using modern symbolic notation, Newton's second law can be written as a vector differential equation:

$$\mathbf{F}_{\text{net}} = \frac{\mathrm{d}(m\mathbf{v})}{\mathrm{d}t} = m\,\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}$$

where **F** is the force vector, *m* is the mass of the body, **v** is the velocity vector and *t* is time.

The product of the mass and velocity is momentum (which Newton himself called "quantity of motion"). Therefore, this equation expresses the physical relationship between force and momentum for a body with constant mass. Because the law describes the motion of bodies of constant mass only[11] [12] [13] , the mass can be moved outside the differential operator. The equation implies that, under zero net force, the momentum of a body is also constant. However, any mass that is gained or lost by the body will cause a change in momentum that is not the result of an external force. This equation does not hold in such cases. See open systems.

It should be noted that, as is consistent with the law of inertia, the time derivative of the momentum is non-zero when the momentum changes *direction*, even if there is no change in its *magnitude*. See time derivative.[14]

By substitution using the definition of acceleration, this differential equation can be rewritten in a more familiar form

$$\mathbf{F} = m\mathbf{a}$$

where

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}.$$

A verbal equivalent of this is "the acceleration of an object is proportional to the force applied, *and inversely proportional to the mass of the object*". In general, at slow speeds (slow relative to the speed of light), the relationship between momentum and velocity is approximately linear. Nearly all speeds within the human experience fall within this category. At higher speeds, however, this approximation becomes increasingly inaccurate and the theory of special relativity must be applied.

Impulse

The term *impulse* is closely related to the second law, and historically speaking is closer to the original meaning of the law.[15] The meaning of an impulse is as follows:[16] [17]

An **impulse** occurs when a force **F** acts over an interval of time Δ*t* and is given by $\int_{\Delta t} \mathbf{F}\,\mathrm{d}t$.

The words *motive force* were used by Newton to describe "impulse" and *motion* to describe momentum; consequently, a historically closer reading of the second law describes the relation between impulse and change of momentum. That is, a mathematical rendering of the original wording resembles a finite difference version of the second law, such as

$$\mathbf{I} = \Delta\mathbf{p} = m\Delta\mathbf{v}$$

where **I** is the impulse, Δ**p** is the change in momentum, *m* is the mass, and Δ**v** is the change in velocity.

The analysis of collisions and impacts uses the impulse concept.[18]

Relativity

Main article: Special Relativity

Open systems

So-called variable mass systems that are not closed systems, like a rocket burning fuel and ejecting spent gases, can not be directly treated by making mass a function of time in the second law.[12] [13] The reasoning, given in *An Introduction to Mechanics* by Kleppner and Kolenkow and other modern texts, is that Newton's second law applies fundamentally to particles. In classical mechanics, particles by definition have constant mass. In case of *well-defined* systems of particles, Newton's law can be extended by summing over all the particles in the system. In this case, we have to refer all vectors to the center of mass. Applying the second law to extended objects implicitly assumes the object to be a well-defined collection of particles. However, 'variable mass' systems like a rocket or a leaking bucket do not consist of a set number of particles. They are not well-defined systems. Therefore Newton's second law can not be applied to them directly.

The general equation of motion for a body whose mass *m* varies with time by either ejecting or accreting mass is obtained by rearranging the second law and adding a term to account for the momentum carried by mass entering or leaving the system,[11]

$$\mathbf{F}_{\text{net}} + \mathbf{u}\frac{\mathrm{d}m}{\mathrm{d}t} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}$$

where **u** is the relative velocity of the escaping or incoming mass with respect to the center of mass of the body. Under some conventions, the quantity **u***d*m*/d*t* on the left-hand side is defined as a force (the force exerted on the body by the changing mass, such as rocket exhaust) and is included in the quantity $\mathbf{F}_{\text{net}}$. Then, by substituting the definition of acceleration, the equation becomes, once again,

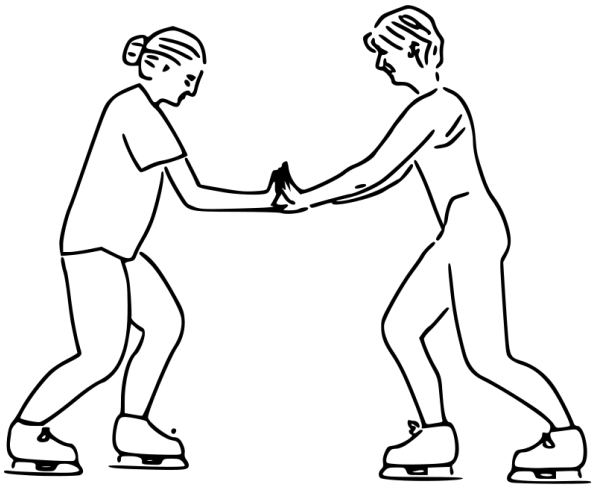
$$\mathbf{F}_{\text{net}} = m\mathbf{a}.$$

Newton's third law: law of reciprocal actions

> *Lex III: Actioni contrariam semper et æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales et in partes contrarias dirigi.*
>
> ***For a force there is always an equal and opposite reaction: or the forces of two bodies on each other are always equal and are directed in opposite directions***.

A more direct translation is:

> *LAW III*: To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts. — Whatever draws or presses another is as much drawn or pressed by that other. If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the stone: for the distended rope, by the same endeavour to relax or unbend itself, will draw the horse as much towards the stone, as it does the stone towards the horse, and will obstruct the progress of the one as much as it advances that of the other. If a body impinges upon another, and by its force changes the motion of the other, that body also (because of the equality of the mutual pressure) will undergo an equal change, in its own motion, toward the contrary part. The changes made by these actions are equal, not in the velocities but in the motions of the bodies; that is to say, if the bodies are not hindered by any other impediments. For, as the motions are equally changed, the changes of the velocities made toward contrary parts are reciprocally proportional to the bodies. This law takes place also in attractions, as will be proved in the next scholium.



Newton's third law. The skaters' forces on each other are equal in magnitude, but act in opposite directions.

In the above, as usual, *motion* is Newton's name for momentum, hence his careful distinction between motion and velocity.

The Third Law means that all forces are *interactions*, and thus that there is no such thing as a unidirectional force. If body *A* exerts a force on body *B*, simultaneously, body *B* exerts a force of the same magnitude body *A*, both forces acting along the same line. As shown in the diagram opposite, the skaters' forces on each other are equal in magnitude, but act in opposite directions. Although the forces are equal, the accelerations are not: the less massive skater will have a greater acceleration due to Newton's second law. It is important to note that the action and reaction act on different objects and do not cancel each other out. The two forces in Newton's third law are of the same type (e.g., if the road exerts a forward frictional force on an accelerating car's tires, then it is also a frictional force that Newton's third law predicts for the tires pushing backward on the road).

Newton used the third law to derive the law of conservation of momentum;[19] however from a deeper perspective, conservation of momentum is the more fundamental idea (derived via Noether's theorem from Galilean invariance), and holds in cases where Newton's third law appears to fail, for instance when force fields as well as particles carry momentum, and in quantum mechanics.

Importance and range of validity

Newton's laws were verified by experiment and observation for over 200 years, and they are excellent approximations at the scales and speeds of everyday life. Newton's laws of motion, together with his law of universal gravitation and the mathematical techniques of calculus, provided for the first time a unified quantitative explanation for a wide range of physical phenomena.

These three laws hold to a good approximation for macroscopic objects under everyday conditions. However, Newton's laws (combined with Universal Gravitation and Classical Electrodynamics) are inappropriate for use in certain circumstances, most notably at very small scales, very high speeds (in special relativity, the Lorentz factor must be included in the expression for momentum along with rest mass and velocity) or very strong gravitational fields. Therefore, the laws cannot be used to explain phenomena such as conduction of electricity in a semiconductor, optical properties of substances, errors in non-relativistically corrected GPS systems and superconductivity. Explanation of these phenomena requires more sophisticated physical theory, including General Relativity and Relativistic Quantum Mechanics.

In quantum mechanics concepts such as force, momentum, and position are defined by linear operators that operate on the quantum state; at speeds that are much lower than the speed of light, Newton's laws are just as exact for these operators as they are for classical objects. At speeds comparable to the speed of light, the second law holds in the original form $\mathbf{F} = \mathrm{d}\mathbf{p}/\mathrm{d}t$, which says that the force is the derivative of the momentum of the object with respect to time, but some of the newer versions of the second law (such as the constant mass approximation above) do not hold at relativistic velocities.

Relationship to the conservation laws

In modern physics, the laws of conservation of momentum, energy, and angular momentum are of more general validity than Newton's laws, since they apply to both light and matter, and to both classical and non-classical physics.

This can be stated simply, "Momentum, energy and angular momentum cannot be created or destroyed."

Because force is the time derivative of momentum, the concept of force is redundant and subordinate to the conservation of momentum, and is not used in fundamental theories (e.g. quantum mechanics, quantum electrodynamics, general relativity, etc.). The standard model explains in detail how the three fundamental forces known as gauge forces originate out of exchange by virtual particles. Other forces such as gravity and fermionic degeneracy pressure also arise from the momentum conservation. Indeed, the conservation of 4-momentum in inertial motion via curved space-time results in what we call gravitational force in general relativity theory. Application of space derivative (which is a momentum operator in quantum mechanics) to overlaping wave functions of pair of fermions (particles

with semi-integer spin) results in shifts of maxima of compound wavefunction away from each other, which is observable as "repulsion" of fermions.

Newton stated the third law within a world-view that assumed instantaneous action at a distance between material particles. However, he was prepared for philosophical criticism of this action at a distance, and it was in this context that he stated the famous phrase "I feign no hypotheses". In modern physics, action at a distance has been completely eliminated, except for subtle effects involving quantum entanglement. However in modern engineering in all practical applications involving the motion of vehicles and satellites, the concept of action at a distance is used extensively.

Conservation of energy was discovered nearly two centuries after Newton's lifetime, the long delay occurring because of the difficulty in understanding the role of microscopic and invisible forms of energy such as heat and infra-red light.

See also

- Scientific laws named after people
- Mercury, orbit of
- Galilean invariance
- Modified Newtonian dynamics
- Lagrangian mechanics
- Hamiltonian mechanics
- Principle of least action
- Euler's laws

References and notes

[1] See the *Principia* on line at Andrew Motte Translation (http://ia310114.us.archive.org/2/items/newtonspmathema00newtrich/newtonspmathema00newtrich.pdf)

[2] Andrew Motte translation of Newton's *Principia* (1687) *Axioms or Laws of Motion* (http://members.tripod.com/~gravitee/axioms.htm)

[3] In the second law, *m* must be treated as the relatistic mass, producing the relativistic expression for momentum, and the third law must be modified to allow for the finite signal propagation speed between distant interacting particles.

[4] Isaac Newton, *The Principia*, A new translation by I.B. Cohen and A. Whitman, University of California press, Berkeley 1999.

[5] NMJ Woodhouse (2003). *http://books.google.com/books?id=ggPXQAeeRLgC&printsec=frontcover&dq=isbn=1852334266#PPA6,M1|Special relativity*. London/Berlin: Springer. p. 6. ISBN 1-85233-426-6. http://books.google.com/books?id=ggPXQAeeRLgC&printsec=frontcover&dq=isbn=1852334266#PPA6,M1.

[6] Galili, I. & Tseitlin, M. (2003), "Newton's first law: text, translations, interpretations, and physics education.", *Science and Education* **12 (1)**: 45–73, doi: 10.1023/A:1022632600805 (http://dx.doi.org/10.1023/A:1022632600805)

[7] On a more technical note, although Newton's laws are not applicable on non-inertial frames of reference, such as the accelerating airplane, they can be made to do so with the introduction of a "fictitious force" acting on the entire system: basically, by introducing a force that quantifies the anomalous motion of objects within that system (such as the ball moving without an apparent influence in the example above

[8] Abdus Salam (1984), "Islam and Science". In C. H. Lai (1987), *Ideals and Realities: Selected Essays of Abdus Salam*, 2nd ed., World Scientific, Singapore, p. 179-213.

[9] Fernando Espinoza (2005). "An analysis of the historical development of ideas about motion and its implications for teaching", *Physics Education* **40** (2), p. 141.

[10] According to Maxwell in *Matter and Motion*, Newton meant by *motion* "the quantity of matter moved as well as the rate at which it travels" and by *impressed force* he meant "the time during which the force acts as well as the intensity of the force". See Harman and Shapiro,cited below.

[11] Plastino, Angel R.; Muzzio, Juan C. (1992). "http://articles.adsabs.harvard.edu//full/1992CeMDA..53..227P/0000227.000.html|On the use and abuse of Newton's second law for variable mass problems" (in English). *Celestial Mechanics and Dynamical Astronomy* (Netherlands: Kluwer Academic Publishers) **vol. 53** (no. 3): pp. 227-232. ISSN 0923-2958 (http://worldcat.org/issn/0923-2958). http://articles.adsabs.harvard.edu//full/1992CeMDA..53..227P/0000227.000.html. Retrieved on 11 June 2009. "We may conclude emphasizing that Newton's second law is valid for constant mass only. When the mass varies due to accretion or ablation, [an alternate equation explicitly accounting for the changing mass] should be used."

[12] Halliday; Resnick. *Physics*. **1**. pp. 199. "It is important to note that we *cannot* derive a general expression for Newton's second law for variable mass systems by treating the mass in **F** = d**P**/dt = d(M**v**) as a *variable*. [...] We *can* use **F** = d**P**/dt to analyze variable mass systems *only* if we apply it to an *entire system of constant mass* having parts among which there is an interchange of mass." [Emphasis as in the original]

[13] Kleppner; Kolenkow. *An Introduction to Mechanics*. pp. 133–134. "Recall that **F** = d**P**/dt was established for a system composed of a certain set of particles...it is essential to deal with the same set of particles throughout the time interval...Consequently, the mass of the system can not change during the time of interest."

[14] The use of algebraic expressions became popular during the 18th century, after Newton's death, while vector notation dates to the late 19th century. The *Principia* expresses mathematical theorems in words and consistently uses geometrical rather than algebraic proofs.

[15] I Bernard Cohen (Peter M. Harman & Alan E. Shapiro, Eds) (2002). *http://books.google.com/books?id=oYZ-0PUrjBcC&pg=PA353&dq=impulse+momentum+"rate+of+change"+-angular+date:200 investigation of difficult things : essays on Newton and the history of the exact sciences in honour of D.T. Whiteside*. Cambridge UK: Cambridge University Press. p. 353. ISBN 052189266X. http://books.google.com/books?id=oYZ-0PUrjBcC&pg=PA353&dq=impulse+momentum+%22rate+of+change%22+-angular+date:2000-2009&lr=&as_brr=0&sig=xM_5Q-nrbPkLLKcXAAbmogvVTcU.

[16] Hannah, J, Hillier, M J, *Applied Mechanics*, p221, Pitman Paperbacks, 1971

[17] Raymond A. Serway, Jerry S. Faughn (2006). *http://books.google.com/books?id=wDKD4IggBJ4C&pg=PA247&dq=impulse+momentum+"rate+of+change"&lr=&as_brr=0&s Physics*. Pacific Grove CA: Thompson-Brooks/Cole. p. 161. ISBN 0534997244. http://books.google.com/books?id=wDKD4IggBJ4C&pg=PA247&dq=impulse+momentum+%22rate+of+change%22&lr=&as_brr=0&sig=Up5LC1E784npQuR2lyDde6SetoQ#PPA161,M1.

[18] WJ Stronge (2004). *http://books.google.com/books?id=nHgcS0bfZ28C&pg=PA12&dq=impulse+momentum+"rate+of+change"+-angular+date:200 mechanics*. Cambridge UK: Cambridge University Press. p. 12 ff. ISBN 0521602890. http://books.google.com/books?id=nHgcS0bfZ28C&pg=PA12&dq=impulse+momentum+%22rate+of+change%22+-angular+date:2000-2009&lr=&as_brr=0&sig=YVDmNVMz38AubS-5lvRADvD2n6k.

[19] Newton, *Principia*, Corollary III to the laws of motion

Further reading

- Marion, Jerry and Thornton, Stephen. *Classical Dynamics of Particles and Systems*. Harcourt College Publishers, 1995. ISBN 0-03-097302-3
- Fowles, G. R. and Cassiday, G. L. *Analytical Mechanics* (6ed). Saunders College Publishing, 1999. ISBN 0-03-022317-2
- Feynman R P, Leighton R B & Sands M (2006). *http://worldcat.org/oclc/61355214&referer=brief_results|The Feynman lectures on physics*. **Vol. 1**. Pearson/Addison-Wesley. ISBN 0805390499. http://worldcat.org/oclc/61355214&referer=brief_results.
- Likins, Peter W. *Elements of Engineering Mechanics* McGraw-Hill Book Company, 1973. ISBN 0-07-037852-5

External links

- MIT Physics video lecture (http://academicearth.org/lectures/newtons-three-laws) on Newton's three laws
- Science aid: Newton's laws of motion (http://www.scienceaid.co.uk/physics/forces/power.html)
- Newtonian Physics (http://www.lightandmatter.com/html_books/1np/ch04/ch04.html) - an on-line textbook
- Motion Mountain (http://www.motionmountain.net) - an on-line textbook (see also Motion Mountain)
- Newtonian attraction for three Planets (http://twt.mpei.ac.ru/MAS/Worksheets/3_Planets.mcd) (Mathcad Application Server)
- Gravity – Newton's Law for Kids (http://www.projectshum.org/Gravity/)
- Simulation on Newton's first law of motion (http://phy.hk/wiki/englishhtm/firstlaw.htm)
- " Newton's Second Law (http://demonstrations.wolfram.com/NewtonsSecondLaw/)" by Enrique Zeleny, Wolfram Demonstrations Project.

Analytical dynamics

In classical mechanics, **analytical dynamics**, or more briefly **dynamics**, is concerned about the relationship between motion of bodies and its causes, namely the forces acting on the bodies and the properties of the bodies (particularly mass and moment of inertia). The foundation of modern day dynamics is Newtonian mechanics and its reformulation as Lagrangian mechanics and Hamiltonian mechanics.[1] [2] The field has a long and important history, as remarked by Hamilton:

> The theoretical development of the laws of motion of bodies is a problem of such interest and importance that it has engaged the attention of all the eminent mathematicians since the invention of the dynamics as a mathematical science by Galileo, and especially since the wonderful extension which was given to that science by Newton
>
> – *William Rowan Hamilton, 1834 (Transcribed in Classical Mechanics by J.R. Taylor, p. 237*[3] *)*

Some authors (for example, Taylor (2005)[3] and Greenwood (1997)[4]) include special relativity within classical dynamics.

Relationship to statics, kinetics, and kinematics

Historically, there were three branches of classical mechanics: "statics" (the study of equilibrium and its relation to forces); "kinetics" (the study of motion and its relation to forces)[5] and "kinematics" (dealing with the implications of observed motions without regard for circumstances causing them).[6] These three subjects have been connected to *dynamics* in several ways. One approach combined statics and kinetics under the name dynamics, which became the branch dealing with determination of the motion of bodies resulting from the action of specified forces[7] ; another approach separated statics, and combined kinetics and kinematics under the rubric dynamics.[8] [9] This approach is common in engineering books on mechanics, and is still in widespread use among

mechanicians.

Fundamental importance in engineering, diminishing emphasis in physics

Today, *dynamics* and *kinematics* continue to be considered the two pillars of classical mechanics. Dynamics is still included in mechanical, aerospace, and other engineering curriculums because of its importance in machine design, the design of land, sea, air, and space vehicles and other applications. However, few modern physicists concern themselves with an independent treatment of "dynamics" or "kinematics", nevermind "statics" or "kinetics". Instead, the entire undifferentiated subject is referred to as *classical mechanics*. In fact, many undergraduate and graduate text books since mid-20th century on "classical mechanics" lack chapters titled "dynamics" or "kinematics" .[3] [10] [11] [12] [13] [14] [15] [16] [17] In these books, although the word "dynamics" is used when acceleration is ascribed to a force, the word "kinetics" is never mentioned. However, clear exceptions exist. Prominent examples include *The Feynman Lectures on Physics*.[18]

Fundamental Principles

- Newton's laws of motion
  - Inertia
  - Acceleration
  - Momentum
  - Reaction
- Newton's law of universal gravitation
- Special theory of relativity

Axioms and mathematical treatments

- Variational principles and Lagrange's equations
- Hamilton's equations
- Canonical transformations
- Hamilton-Jacobi Theory

Related engineering branches

- Particle dynamics
- Rigid body dynamics
- Soft body dynamics
- Fluid dynamics
  - Hydrodynamics
  - Gas dynamics
  - Aerodynamics

Related subjects

- Statics

References

[1] Chris Doran, Anthony N. Lasenby (2003). *http://books.google.com/books?id=VW4yt0WHdjoC&pg=PA54&dq=classical+dynamics+-quantum+date:2002-2009&lr=&as_br*
Algebra for Physicists. Cambridge University Press. p. p. 54. ISBN 0521480221. http://books.google.com/books?id=VW4yt0WHdjoC&pg=PA54&dq=classical+dynamics+-quantum+date:2002-2009&lr=&as_brr=0&sig=ACfU3U11syztEgIW0cnsvMxQhO1nQ51KRw.

[2] Cornelius Lanczos (1986). *http://books.google.com/books?id=ZWoYYr8wk2IC&pg=PR4&dq=isbn=0486650677&sig=ACfU3U2R5sLjGS22S-h8Z_j9RiPJnK*
variational principles of mechanics (Reprint of 4th Edition of 1970 ed.). Dover Publications Inc.. p. pp. 5-6. ISBN 0-486-65067-7. http://books.google.com/books?id=ZWoYYr8wk2IC&pg=PR4&dq=isbn=0486650677&sig=ACfU3U2R5sLjGS22S-h8Z_j9RiPJnKcKZg#PPA5,M1.

[3] John Robert Taylor (2005). *http://books.google.com/books?id=P1kCtNr-pJsC&q=dynamics#search|Classical Mechanics*. University Science Books. ISBN 189138922X, 9781891389221. http://books.google.com/books?id=P1kCtNr-pJsC&q=dynamics#search.

[4] Donald T Greenwood (1997). *http://books.google.com/books?id=x7rj83I98yMC&printsec=frontcover&dq=classical+dynamics&lr=&as_brr=0&sig=ACfU3U2*
Mechanics (Reprint of 1977 ed.). Courier Dover Publications. p. p. 1. ISBN 0486696901. http://books.google.com/books?id=x7rj83I98yMC&printsec=frontcover&dq=classical+dynamics&lr=&as_brr=0&sig=ACfU3U2-b1lzGZZqchuPz0_7Pu7IF-5UyQ#PPA1,M1.

[5] Thomas Wallace Wright (1896). *http://books.google.com/books?id=-LwLAAAAYAAJ&printsec=frontcover&dq=mechanics+kinetics&lr=&as_brr=0#PPA85,M1|E*
of Mechanics Including Kinematics, Kinetics and Statics: with applications. E. and F. N. Spon. p. p. 85. http://books.google.com/books?id=-LwLAAAAYAAJ&printsec=frontcover&dq=mechanics+kinetics&lr=&as_brr=0#PPA85,M1.

[6] Edmund Taylor Whittaker (1988). *http://books.google.com/books?id=epH1hCB7N2MC&printsec=frontcover&dq=inauthor:"E+T+Whittaker"&lr=&as_brr=0&sig*
Treatise on the Analytical Dynamics of Particles and Rigid Bodies: With an Introduction to the Problem of Three Bodies (Fourth edition of 1936 with foreword by Sir William McCrea ed.). Cambridge University Press. p. Chapter 1, p. 1. ISBN 0521358833. http://books.google.com/books?id=epH1hCB7N2MC&printsec=frontcover&dq=inauthor:%22E+T+Whittaker%22&lr=&as_brr=0&sig=SN7_oYmNYM4QRSgjULXBU5jeQrA&source=gbs_book_other_versions_r&cad=0_2#PPA1,M1.

[7] James Gordon MacGregor (1887). *http://books.google.com/books?id=3yMQAAAAYAAJ&printsec=frontcover&dq=kinematics+dynamics&lr=&as_brr=0#PPR5,M1*
Elementary Treatise on Kinematics and Dynamics. Macmillan. p. p. *v*. http://books.google.com/books?id=3yMQAAAAYAAJ&printsec=frontcover&dq=kinematics+dynamics&lr=&as_brr=0#PPR5,M1.

[8] Stephen Timoshenko, Donovan Harold Young (1956). *http://books.google.ca/books?id=I548AAAAIAAJ&q=engineering+mechanics+inauthor:Timoshenko&dq=engineering+mechanic*
mechanics. McGraw Hill. http://books.google.ca/books?id=I548AAAAIAAJ&q=engineering+mechanics+inauthor:Timoshenko&dq=engineering+mechanics+inauthor:Timoshenko&lr=&as_brr=0&ei=hX_SSO2_E432sgPuzrGTBw&pgis=1.

[9] Lakshmana C. Rao, J. Lakshminarasimhan, Raju Sethuraman, Srinivasan M. Sivakumar (2004). *http://books.google.com/books?id=F7gaa1ShPKIC&printsec=frontcover&dq=statics+dynamics&lr=&as_brr=0&sig=ACfU3U2h*
mechanics. PHI Learning Pvt. Ltd.. p. p. *vi*. ISBN 8120321898. http://books.google.com/books?id=F7gaa1ShPKIC&printsec=frontcover&dq=statics+dynamics&lr=&as_brr=0&sig=ACfU3U2haQ0TLc90YwYiTtuhvIgfA6ZXEQ#PPR6,M1.

[10] David Hestenes (1999). *http://books.google.com/books?id=eU2qm8wavRwC&pg=PA198&dq=dynamics+kinematics&lr=&as_brr=0&sig=ACfU3U2C3a*
Foundations for Classical Mechanics. Springer. p. p. 198. ISBN 0792355148. http://books.google.com/books?id=eU2qm8wavRwC&pg=PA198&dq=dynamics+kinematics&lr=&as_brr=0&sig=ACfU3U2C3aW_zumPCy2Doe4K4NKsjJXKeQ.

[11] R. Douglas Gregory (2006). *http://books.google.com/books?id=uAfUQmQbzOkC&q=dynamics#search|Classical Mechanics: An Undergraduate Text*. Cambridge University Press. ISBN 0521826780, 9780521826785. http://books.google.com/books?id=uAfUQmQbzOkC&q=dynamics#search.

[12] Landau, L. D.; Lifshitz, E. M.; Sykes, J.B.; Bell, J. S. (1976), *http://books.google.com/books?id=LmAV8q_OOOgC|Mechanics*, **1**, Butterworth-Heinemann, ISBN 0750628960, 9780750628969, http://books.google.com/books?id=LmAV8q_OOOgC

[13] Jorge Valenzuela José, Eugene Jerome Saletan (1998). *http://books.google.com/books?id=ZW0L5Xe9zhwC|Classical Dynamics: A Contemporary Approach*. Cambridge University Press. ISBN 0521636361, 9780521636360. http://books.google.com/books?id=ZW0L5Xe9zhwC.

[14] T. W. B. Kibble, Frank H. Berkshire (2004). *http://books.google.com/books?id=0a8dk0eDxgEC|Classical Mechanics*. Imperial College Press. ISBN 1860944353, 9781860944352. http://books.google.com/books?id=0a8dk0eDxgEC.

[15] Walter Greiner, S. Allan Bromley (2003). *http://books.google.com/books?id=L_APSPGoI5sC|Classical Mechanics: Point Particles and Relativity*. Springer. ISBN 0387955860, 9780387955865. http://books.google.com/books?id=L_APSPGoI5sC.

[16] Gerald Jay Sussman, Jack Wisdom Meinhard, Edwin Mayer (2001). *http://books.google.com/books?client=firefox-a&id=H_6Ux04cPv8C&q=dynamics#search|Structure and Interpretation of Classical Mechanics*. MIT Press. ISBN 0262194554, 9780262194556. http://books.google.com/books?client=firefox-a&id=H_6Ux04cPv8C&q=dynamics#search.

[17] Harald Iro (2002). *http://books.google.com/books?id=-L5ckgdxA5YC|A Modern Approach to Classical Mechanics*. World Scientific. ISBN 9812382135, 9789812382139. http://books.google.com/books?id=-L5ckgdxA5YC.

[18] Feynman, RP; Leighton, RB; Sands, M (2003), *The Feynman Lectures on Physics*, **Vol. 1** (Reprint of 1963 lectures ed.), Perseus Books Group, p. Ch. 9 *Newton's Laws of Dynamics*, ISBN 0738209309

Molecular dynamics

Molecular dynamics (**MD**) is a form of computer simulation in which atoms and molecules are allowed to interact for a period of time by approximations of known physics, giving a view of the motion of the atoms. Because molecular systems generally consist of a vast number of particles, it is impossible to find the properties of such complex systems analytically. When the number of bodies are more than two no analytical solutions can be found and result in chaotic motion (see n-body problem). MD simulation circumvents this problem by using numerical methods. It represents an interface between laboratory experiments and theory, and can be understood as a "virtual experiment". MD probes the relationship between molecular structure, movement and function. Molecular dynamics is a multidisciplinary method. Its laws and theories stem from mathematics, physics, and chemistry, and it employs algorithms from computer science and information theory. It was originally conceived within theoretical physics in the late 1950s[1] and early 1960s [2] , but is applied today mostly in materials science and modeling of biomolecules.

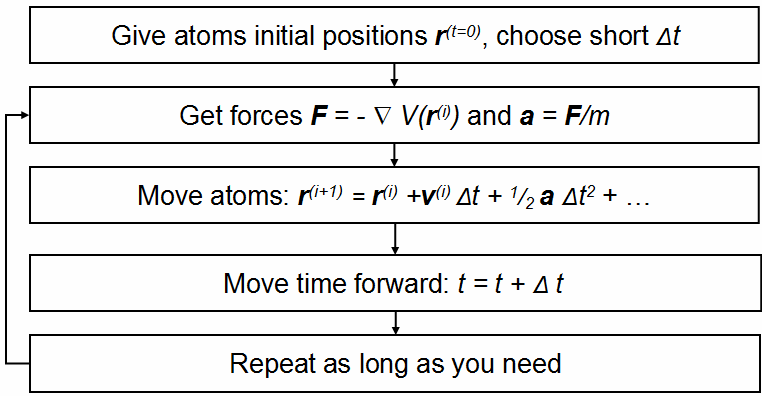
Before it became possible to simulate molecular dynamics with computers, some undertook the hard work of trying it with physical models such as macroscopic spheres. The idea was to arrange them to replicate the properties of a liquid. J.D. Bernal said, in 1962: "... I took a number of rubber balls and stuck them together with rods of a selection of different lengths ranging from 2.75 to 4 inches. I tried to do this in the first place as casually as possible, working in my own office, being interrupted every five minutes or so and not remembering what I had done before the interruption."[3] Fortunately, now computers keep track of bonds during a simulation.

Molecular dynamics is a specialized discipline of molecular modeling and computer simulation based on statistical mechanics; the main justification of the MD method is that statistical ensemble averages are equal to time averages of the system, known as the ergodic hypothesis. MD has also been termed "statistical mechanics by numbers" and "Laplace's vision of Newtonian mechanics" of predicting the future by animating nature's

forces[4] [5] and allowing insight into molecular motion on an atomic scale. However, long MD simulations are mathematically ill-conditioned, generating cumulative errors in numerical integration that can be minimized with proper selection of algorithms and parameters, but not eliminated entirely. Furthermore, current potential functions are, in many cases, not sufficiently accurate to reproduce the dynamics of molecular systems, so the much more computationally demanding Ab Initio Molecular Dynamics method must be used. Nevertheless, molecular dynamics techniques allow detailed time and space resolution into representative behavior in phase space.

Areas of Application

There is a significant difference between the focus and methods used by chemists and physicists, and this is reflected in differences in the jargon used by the different fields. In chemistry and biophysics, the interaction between the particles is either described by a "force field" (**classical MD**), a quantum chemical model, or a mix between the two. These terms are not used in physics, where the interactions are usually described by the name of the theory or approximation being used and called the potential energy, or just "potential".



Highly simplified description of the molecular dynamics simulation algorithm. The simulation proceeds iteratively by alternatively calculating forces and solving the equations of motion based on the accelerations obtained from the new forces. In practise, almost all MD codes use much more complicated versions of the algorithm, including two steps (predictor and corrector) in solving the equations of motion and many additional steps for e.g. temperature and pressure control, analysis and output.

Beginning in theoretical physics, the method of MD gained popularity in materials science and since the 1970s also in biochemistry and biophysics. In chemistry, MD serves as an important tool in protein structure determination and refinement using experimental tools such as X-ray crystallography and NMR. It has also been applied with limited success as a method of refining protein structure predictions. In physics, MD is used to examine the dynamics of atomic-level phenomena that cannot be observed directly, such as thin film growth and ion-subplantation. It is also used to examine the physical properties of nanotechnological devices that have not or cannot yet be created.

In applied mathematics and theoretical physics, molecular dynamics is a part of the research realm of dynamical systems, ergodic theory and statistical mechanics in general. The concepts of energy conservation and molecular entropy come from thermodynamics. Some techniques to calculate conformational entropy such as principal components analysis come from information theory. Mathematical techniques such as the transfer operator become applicable when MD is seen as a Markov chain. Also, there is a large community of mathematicians working on volume preserving, symplectic integrators for more computationally efficient MD simulations.

MD can also be seen as a special case of the discrete element method (DEM) in which the particles have spherical shape (e.g. with the size of their van der Waals radii.) Some authors in the DEM community employ the term MD rather loosely, even when their simulations do not model actual molecules.

Design Constraints

Design of a molecular dynamics simulation should account for the available computational power. Simulation size (n=number of particles), timestep and total time duration must be selected so that the calculation can finish within a reasonable time period. However, the simulations should be long enough to be relevant to the time scales of the natural processes being studied. To make statistically valid conclusions from the simulations, the time span simulated should match the kinetics of the natural process. Otherwise, it is analogous to making conclusions about how a human walks from less than one footstep. Most scientific publications about the dynamics of proteins and DNA use data from simulations spanning nanoseconds (1E-9 s) to microseconds (1E-6 s). To obtain these simulations, several CPU-days to CPU-years are needed. Parallel algorithms allow the load to be distributed among CPUs; an example is the spatial decomposition in LAMMPS.

During a classical MD simulation, the most CPU intensive task is the evaluation of the potential (force field) as a function of the particles' internal coordinates. Within that energy evaluation, the most expensive one is the non-bonded or non-covalent part. In Big O notation, common molecular dynamics simulations scale by $O(n^2)$if all pair-wise electrostatic and van der Waals interactions must be accounted for explicitly. This computational cost can be reduced by employing electrostatics methods such as Particle Mesh Ewald ($O(n \log(n))$) or good spherical cutoff techniques ($O(n)$).

Another factor that impacts total CPU time required by a simulation is the size of the integration timestep. This is the time length between evaluations of the potential. The timestep must be chosen small enough to avoid discretization errors (i.e. smaller than the fastest vibrational frequency in the system). Typical timesteps for classical MD are in the order of 1 femtosecond (1E-15 s). This value may be extended by using algorithms such as SHAKE, which fix the vibrations of the fastest atoms (e.g. hydrogens) into place. Multiple time scale methods have also been developed, which allow for extended times between updates of slower long-range forces.[6] [7] [8]

For simulating molecules in a solvent, a choice should be made between explicit solvent and implicit solvent. Explicit solvent particles (such as the TIP3P and SPC/E water models) must be calculated expensively by the force field, while implicit solvents use a mean-field approach. Using an explicit solvent is computationally expensive, requiring inclusion of about ten times more particles in the simulation. But the granularity and viscosity of explicit solvent is essential to reproduce certain properties of the solute molecules. This is especially important to reproduce kinetics.

In all kinds of molecular dynamics simulations, the simulation box size must be large enough to avoid boundary condition artifacts. Boundary conditions are often treated by choosing fixed values at the edges, or by employing periodic boundary conditions in which one side of the simulation loops back to the opposite side, mimicking a bulk phase.

Microcanonical ensemble (NVE)

In the **microcanonical**, or **NVE** ensemble, the system is isolated from changes in moles (N), volume (V) and energy (E). It corresponds to an adiabatic process with no heat exchange. A microcanonical molecular dynamics trajectory may be seen as an exchange of potential and kinetic energy, with total energy being conserved. For a system of N particles with coordinates X and velocities V, the following pair of first order differential equations may be written in Newton's notation as

$$F(X) = -\nabla U(X) = M\dot{V}(t)$$
$$V(t) = \dot{X}(t).$$

The potential energy function $U(X)$ of the system is a function of the particle coordinates X. It is referred to simply as the "potential" in Physics, or the "force field" in Chemistry. The first equation comes from Newton's laws; the force F acting on each particle in the system can be calculated as the negative gradient of $U(X)$.

For every timestep, each particle's position X and velocity V may be integrated with a symplectic method such as Verlet. The time evolution of X and V is called a trajectory. Given the initial positions (e.g. from theoretical knowledge) and velocities (e.g. randomized Gaussian), we can calculate all future (or past) positions and velocities.

One frequent source of confusion is the meaning of temperature in MD. Commonly we have experience with macroscopic temperatures, which involve a huge number of particles. But temperature is a statistical quantity. If there is a large enough number of atoms, statistical temperature can be estimated from the *instantaneous temperature*, which is found by equating the kinetic energy of the system to $nk_BT/2$ where n is the number of degrees of freedom of the system.

A temperature-related phenomenon arises due to the small number of atoms that are used in MD simulations. For example, consider simulating the growth of a copper film starting with a substrate containing 500 atoms and a deposition energy of 100 eV. In the real world, the 100 eV from the deposited atom would rapidly be transported through and shared among a large number of atoms (10^{10} or more) with no big change in temperature. When there are only 500 atoms, however, the substrate is almost immediately vaporized by the deposition. Something similar happens in biophysical simulations. The temperature of the system in NVE is naturally raised when macromolecules such as proteins undergo exothermic conformational changes and binding.

Canonical ensemble (NVT)

In the canonical ensemble, moles (N), volume (V) and temperature (T) are conserved. It is also sometimes called constant temperature molecular dynamics (CTMD). In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat.

A variety of thermostat methods are available to add and remove energy from the boundaries of an MD system in a realistic way, approximating the canonical ensemble. Popular techniques to control temperature include the Nosé-Hoover thermostat, the Berendsen thermostat, and Langevin dynamics. Note that the Berendsen thermostat might introduce the flying ice cube effect, which leads to unphysical translations and rotations of the simulated system.

Isothermal-Isobaric (NPT) ensemble

In the isothermal-isobaric ensemble, moles (N), pressure (P) and temperature (T) are conserved. In addition to a thermostat, a barostat is needed. It corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure.

In the simulation of biological membranes, isotropic pressure control is not appropriate. For lipid bilayers, pressure control occurs under constant membrane area (NPAT) or constant surface tension "gamma" (NPγT).

Generalized ensembles

The replica exchange method is a generalized ensemble. It was originally created to deal with the slow dynamics of disordered spin systems. It is also called parallel tempering. The replica exchange MD (REMD) formulation [9] tries to overcome the multiple-minima problem by exchanging the temperature of non-interacting replicas of the system running at several temperatures.

Potentials in MD simulations

A molecular dynamics simulation requires the definition of a potential function, or a description of the terms by which the particles in the simulation will interact. In chemistry and biology this is usually referred to as a force field. Potentials may be defined at many levels of physical accuracy; those most commonly used in chemistry are based on molecular mechanics and embody a classical treatment of particle-particle interactions that can reproduce structural and conformational changes but usually cannot reproduce chemical reactions.

The reduction from a fully quantum description to a classical potential entails two main approximations. The first one is the Born-Oppenheimer approximation, which states that the dynamics of electrons is so fast that they can be considered to react instantaneously to the motion of their nuclei. As a consequence, they may be treated separately. The second one treats the nuclei, which are much heavier than electrons, as point particles that follow classical Newtonian dynamics. In classical molecular dynamics the effect of the electrons is approximated as a single potential energy surface, usually representing the ground state.

When finer levels of detail are required, potentials based on quantum mechanics are used; some techniques attempt to create hybrid classical/quantum potentials where the bulk of the system is treated classically but a small region is treated as a quantum system, usually undergoing a chemical transformation.

Empirical potentials

Empirical potentials used in chemistry are frequently called force fields, while those used in materials physics are called just empirical or analytical potentials.

Most force fields in chemistry are empirical and consist of a summation of bonded forces associated with chemical bonds, bond angles, and bond dihedrals, and non-bonded forces associated with van der Waals forces and electrostatic charge. Empirical potentials represent quantum-mechanical effects in a limited way through ad-hoc functional approximations. These potentials contain free parameters such as atomic charge, van der Waals parameters reflecting estimates of atomic radius, and equilibrium bond length, angle, and dihedral; these are obtained by fitting against detailed electronic calculations

(quantum chemical simulations) or experimental physical properties such as elastic constants, lattice parameters and spectroscopic measurements.

Because of the non-local nature of non-bonded interactions, they involve at least weak interactions between all particles in the system. Its calculation is normally the bottleneck in the speed of MD simulations. To lower the computational cost, force fields employ numerical approximations such as shifted cutoff radii, reaction field algorithms, particle mesh Ewald summation, or the newer Particle-Particle Particle Mesh (P3M).

Chemistry force fields commonly employ preset bonding arrangements (an exception being *ab-initio* dynamics), and thus are unable to model the process of chemical bond breaking and reactions explicitly. On the other hand, many of the potentials used in physics, such as those based on the bond order formalism can describe several different coordinations of a system and bond breaking. Examples of such potentials include the Brenner potential[10] for hydrocarbons and its further developments for the C-Si-H and C-O-H systems. The ReaxFF potential[11] can be considered a fully reactive hybrid between bond order potentials and chemistry force fields.

Pair potentials vs. many-body potentials

The potential functions representing the non-bonded energy are formulated as a sum over interactions between the particles of the system. The simplest choice, employed in many popular force fields, is the "pair potential", in which the total potential energy can be calculated from the sum of energy contributions between pairs of atoms. An example of such a pair potential is the non-bonded Lennard-Jones potential (also known as the 6-12 potential), used for calculating van der Waals forces.

$$U(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

Another example is the Born (ionic) model of the ionic lattice. The first term in the next equation is Coulomb's law for a pair of ions, the second term is the short-range repulsion explained by Pauli's exclusion principle and the final term is the dispersion interaction term. Usually, a simulation only includes the dipolar term, although sometimes the quadrupolar term is included as well.

$$U_{ij}(r_{ij}) = \sum \frac{z_i z_j}{4\pi\epsilon_0} \frac{1}{r_{ij}} + \sum A_l \exp \frac{-r_{ij}}{p_l} + \sum C_l r_{ij}^{-n_j} + \cdots$$

In many-body potentials, the potential energy includes the effects of three or more particles interacting with each other. In simulations with pairwise potentials, global interactions in the system also exist, but they occur only through pairwise terms. In many-body potentials, the potential energy cannot be found by a sum over pairs of atoms, as these interactions are calculated explicitly as a combination of higher-order terms. In the statistical view, the dependency between the variables cannot in general be expressed using only pairwise products of the degrees of freedom. For example, the Tersoff potential[12] , which was originally used to simulate carbon, silicon and germanium and has since been used for a wide range of other materials, involves a sum over groups of three atoms, with the angles between the atoms being an important factor in the potential. Other examples are the embedded-atom method (EAM)[13] and the Tight-Binding Second Moment Approximation (TBSMA) potentials[14] , where the electron density of states in the region of an atom is calculated from a sum of contributions from surrounding atoms, and the potential energy contribution is then a function of this sum.

Semi-empirical potentials

Semi-empirical potentials make use of the matrix representation from quantum mechanics. However, the values of the matrix elements are found through empirical formulae that estimate the degree of overlap of specific atomic orbitals. The matrix is then diagonalized to determine the occupancy of the different atomic orbitals, and empirical formulae are used once again to determine the energy contributions of the orbitals.

There are a wide variety of semi-empirical potentials, known as tight-binding potentials, which vary according to the atoms being modeled.

Polarizable potentials

Most classical force fields implicitly include the effect of polarizability, e.g. by scaling up the partial charges obtained from quantum chemical calculations. These partial charges are stationary with respect to the mass of the atom. But molecular dynamics simulations can explicitly model polarizability with the introduction of induced dipoles through different methods, such as Drude particles or fluctuating charges. This allows for a dynamic redistribution of charge between atoms which responds to the local chemical environment.

For many years, polarizable MD simulations have been touted as the next generation. For homogenous liquids such as water, increased accuracy has been achieved through the inclusion of polarizability.[15] Some promising results have also been achieved for proteins.[16] However, it is still uncertain how to best approximate polarizability in a simulation.

Ab-initio methods

In classical molecular dynamics, a single potential energy surface (usually the ground state) is represented in the force field. This is a consequence of the Born-Oppenheimer approximation. If excited states, chemical reactions or a more accurate representation is needed, electronic behavior can be obtained from first principles by using a quantum mechanical method, such as Density Functional Theory. This is known as Ab Initio Molecular Dynamics (AIMD). Due to the cost of treating the electronic degrees of freedom, the computational cost of this simulations is much higher than classical molecular dynamics. This implies that AIMD is limited to smaller systems and shorter periods of time.

Ab-initio quantum-mechanical methods may be used to calculate the potential energy of a system on the fly, as needed for conformations in a trajectory. This calculation is usually made in the close neighborhood of the reaction coordinate. Although various approximations may be used, these are based on theoretical considerations, not on empirical fitting. *Ab-Initio* calculations produce a vast amount of information that is not available from empirical methods, such as density of electronic states or other electronic properties. A significant advantage of using *ab-initio* methods is the ability to study reactions that involve breaking or formation of covalent bonds, which correspond to multiple electronic states.

A popular software for *ab-initio* molecular dynamics is the Car-Parrinello Molecular Dynamics (CPMD) package based on the density functional theory.

Hybrid QM/MM

QM (quantum-mechanical) methods are very powerful. However, they are computationally expensive, while the MM (classical or molecular mechanics) methods are fast but suffer from several limitations (require extensive parameterization; energy estimates obtained are not very accurate; cannot be used to simulate reactions where covalent bonds are broken/formed; and are limited in their abilities for providing accurate details regarding the chemical environment). A new class of method has emerged that combines the good points of QM (accuracy) and MM (speed) calculations. These methods are known as mixed or hybrid quantum-mechanical and molecular mechanics methods (hybrid QM/MM). The methodology for such techniques was introduced by Warshel and coworkers. In the recent years have been pioneered by several groups including: Arieh Warshel (University of Southern California), Weitao Yang (Duke University), Sharon Hammes-Schiffer (The Pennsylvania State University), Donald Truhlar and Jiali Gao (University of Minnesota) and Kenneth Merz (University of Florida).

The most important advantage of hybrid QM/MM methods is the speed. The cost of doing classical molecular dynamics (MM) in the most straightforward case scales $O(n^2)$, where N is the number of atoms in the system. This is mainly due to electrostatic interactions term (every particle interacts with every other particle). However, use of cutoff radius, periodic pair-list updates and more recently the variations of the particle-mesh Ewald's (PME) method has reduced this between O(N) to $O(n^2)$. In other words, if a system with twice many atoms is simulated then it would take between twice to four times as much computing power. On the other hand the simplest *ab-initio* calculations typically scale $O(n^3)$ or worse (Restricted Hartree-Fock calculations have been suggested to scale $\sim O(n^{2.7})$). To overcome the limitation, a small part of the system is treated quantum-mechanically (typically active-site of an enzyme) and the remaining system is treated classically.

In more sophisticated implementations, QM/MM methods exist to treat both light nuclei susceptible to quantum effects (such as hydrogens) and electronic states. This allows generation of hydrogen wave-functions (similar to electronic wave-functions). This methodology has been useful in investigating phenomenon such as hydrogen tunneling. One example where QM/MM methods have provided new discoveries is the calculation of hydride transfer in the enzyme liver alcohol dehydrogenase. In this case, tunneling is important for the hydrogen, as it determines the reaction rate.[17]

Coarse-graining and reduced representations

At the other end of the detail scale are coarse-grained and lattice models. Instead of explicitly representing every atom of the system, one uses "pseudo-atoms" to represent groups of atoms. MD simulations on very large systems may require such large computer resources that they cannot easily be studied by traditional all-atom methods. Similarly, simulations of processes on long timescales (beyond about 1 microsecond) are prohibitively expensive, because they require so many timesteps. In these cases, one can sometimes tackle the problem by using reduced representations, which are also called coarse-grained models.

Examples for coarse graining (CG) methods are discontinuous molecular dynamics (CG-DMD)[18] [19] and Go-models[20] . Coarse-graining is done sometimes taking larger pseudo-atoms. Such united atom approximations have been used in MD simulations of biological membranes. The aliphatic tails of lipids are represented by a few pseudo-atoms

by gathering 2-4 methylene groups into each pseudo-atom.

The parameterization of these very coarse-grained models must be done empirically, by matching the behavior of the model to appropriate experimental data or all-atom simulations. Ideally, these parameters should account for both enthalpic and entropic contributions to free energy in an implicit way. When coarse-graining is done at higher levels, the accuracy of the dynamic description may be less reliable. But very coarse-grained models have been used successfully to examine a wide range of questions in structural biology.

Examples of applications of coarse-graining in biophysics:

- protein folding studies are often carried out using a single (or a few) pseudo-atoms per amino acid;
- DNA supercoiling has been investigated using 1-3 pseudo-atoms per basepair, and at even lower resolution;
- Packaging of double-helical DNA into bacteriophage has been investigated with models where one pseudo-atom represents one turn (about 10 basepairs) of the double helix;
- RNA structure in the ribosome and other large systems has been modeled with one pseudo-atom per nucleotide.

The simplest form of coarse-graining is the "united atom" (sometimes called "extended atom") and was used in most early MD simulations of proteins, lipids and nucleic acids. For example, instead of treating all four atoms of a CH_3 methyl group explicitly (or all three atoms of CH_2 methylene group), one represents the whole group with a single pseudo-atom. This pseudo-atom must, of course, be properly parameterized so that its van der Waals interactions with other groups have the proper distance-dependence. Similar considerations apply to the bonds, angles, and torsions in which the pseudo-atom participates. In this kind of united atom representation, one typically eliminates all explicit hydrogen atoms except those that have the capability to participate in hydrogen bonds ("polar hydrogens"). An example of this is the Charmm 19 force-field.

The polar hydrogens are usually retained in the model, because proper treatment of hydrogen bonds requires a reasonably accurate description of the directionality and the electrostatic interactions between the donor and acceptor groups. A hydroxyl group, for example, can be both a hydrogen bond donor and a hydrogen bond acceptor, and it would be impossible to treat this with a single OH pseudo-atom. Note that about half the atoms in a protein or nucleic acid are nonpolar hydrogens, so the use of united atoms can provide a substantial savings in computer time.

Examples of applications

Molecular dynamics is used in many fields of science.

- First macromolecular MD simulation published (1977, Size: 500 atoms, Simulation Time: 9.2 ps=0.0092 ns, Program: CHARMM precursor) Protein: Bovine Pancreatic Trypsine Inhibitor. This is one of the best studied proteins in terms of folding and kinetics. Its simulation published in Nature magazine paved the way for understanding protein motion as essential in function and not just accessory.[21]
- MD is the standard method to treat collision cascades in the heat spike regime, i.e. the effects that energetic neutron and ion irradiation have on solids an solid surfaces.[22] [23]

The following two biophysical examples are not run-of-the-mill MD simulations. They illustrate almost heroic efforts to produce simulations of a system of very large size (a complete virus) and very long simulation times (500 microseconds):

- MD simulation of the complete satellite tobacco mosaic virus (**STMV**) (2006, Size: 1 million atoms, Simulation time: 50 ns, program: NAMD) This virus is a small, icosahedral plant virus which worsens the symptoms of infection by Tobacco Mosaic Virus (TMV). Molecular dynamics simulations were used to probe the mechanisms of viral assembly. The entire STMV particle consists of 60 identical copies of a single protein that make up the viral capsid (coating), and a 1063 nucleotide single stranded RNA genome. One key finding is that the capsid is very unstable when there is no RNA inside. The simulation would take a single 2006 desktop computer around 35 years to complete. It was thus done in many processors in parallel with continuous communication between them.[24]
- Folding Simulations of the Villin Headpiece in All-Atom Detail (2006, Size: 20,000 atoms; Simulation time: 500 μs = 500,000 ns, Program: folding@home) This simulation was run in 200,000 CPU's of participating personal computers around the world. These computers had the folding@home program installed, a large-scale distributed computing effort coordinated by Vijay Pande at Stanford University. The kinetic properties of the Villin Headpiece protein were probed by using many independent, short trajectories run by CPU's without continuous real-time communication. One technique employed was the Pfold value analysis, which measures the probability of folding before unfolding of a specific starting conformation. Pfold gives information about transition state structures and an ordering of conformations along the folding pathway. Each trajectory in a Pfold calculation can be relatively short, but many independent trajectories are needed.[25]

Molecular dynamics algorithms

Integrators

- Verlet-Stoermer integration
- Runge-Kutta integration
- Beeman's algorithm
- Gear predictor - corrector
- Constraint algorithms (for constrained systems)
- Symplectic integrator

Short-range interaction algorithms

- Cell lists
- Verlet list
- Bonded interactions

Long-range interaction algorithms

- Ewald summation
- Particle Mesh Ewald (PME)
- Particle-Particle Particle Mesh P3M
- Reaction Field Method

Parallelization strategies

- Domain decomposition method (Distribution of system data for parallel computing)
- Molecular Dynamics - Parallel Algorithms [26]

Major software for MD simulations

- Abalone (classical, implicit water)
- ABINIT (DFT)
- ACEMD [3] (running on NVIDIA GPUs: heavily optimized with CUDA)
- ADUN [27] (classical, P2P database for simulations)
- AMBER (classical)
- Ascalaph [28] (classical, GPU accelerated)
- CASTEP (DFT)
- CPMD (DFT)
- CP2K [29] (DFT)
- CHARMM (classical, the pioneer in MD simulation, extensive analysis tools)
- COSMOS [11] (classical and hybrid QM/MM, quantum-mechanical atomic charges with BPT)
- Desmond [30] (classical, parallelization with up to thousands of CPU's)
- DL_POLY [31] (classical)
- ESPResSo (classical, coarse-grained, parallel, extensible)
- Fireball [32] (tight-binding DFT)
- GROMACS (classical)
- GROMOS (classical)
- GULP (classical)
- Hippo [33] (classical)
- LAMMPS (classical, large-scale with spatial-decomposition of simulation domain for parallelism)
- MDynaMix (classical, parallel)
- MOLDY [25] (classical, parallel) latest release [34]
- Materials Studio [17] (Forcite MD using COMPASS, Dreiding, Universal, cvff and pcff forcefields in serial or parallel, QMERA (QM+MD), ONESTEP (DFT), etc.)
- MOSCITO (classical)
- NAMD (classical, parallelization with up to thousands of CPU's)
- NEWTON-X [35] (ab initio, surface-hopping dynamics)
- ProtoMol [36] (classical, extensible, includes multigrid electrostatics)
- PWscf (DFT)
- S/PHI/nX [37] (DFT)
- SIESTA (DFT)
- VASP (DFT)
- TINKER (classical)
- YASARA [38] (classical)
- ORAC [39] (classical)
- XMD (classical)

Related software

- VMD - MD simulation trajectories can be visualized and analyzed.
- PyMol - Molecular Visualization software written in python
- Packmol [28] Package for building starting configurations for MD in an automated fashion
- Sirius - Molecular modeling, analysis and visualization of MD trajectories
- esra [40] - Lightweight molecular modeling and analysis library (Java/Jython/Mathematica).
- Molecular Workbench [41] - Interactive molecular dynamics simulations on your desktop
- BOSS - MC in OPLS

Specialized hardware for MD simulations

- Anton - A specialized, massively parallel supercomputer designed to execute MD simulations.
- MDGRAPE - A special purpose system built for molecular dynamics simulations, especially protein structure prediction.

See also

- Molecular graphics
- Molecular modeling
- Computational chemistry
- Energy drift
- Force field in Chemistry
- Force field implementation
- Monte Carlo method
- Molecular Design software
- Molecular mechanics
- Molecular modeling on GPU
- Protein dynamics
- Implicit solvation
- Car-Parrinello method
- Symplectic numerical integration
- Software for molecular mechanics modeling
- Dynamical systems
- Theoretical chemistry
- Statistical mechanics
- Quantum chemistry
- Discrete element method
- List of nucleic acid simulation software

References

[1] Alder, B. J.; T. E. Wainwright (1959). "Studies in Molecular Dynamics. I. General Method". *J. Chem. Phys.* **31** (2): 459. doi: 10.1063/1.1730376 (http://dx.doi.org/10.1063/1.1730376).

[2] A. Rahman (1964). "Correlations in the Motion of Atoms in Liquid Argon". *Phys Rev* **136**: A405-A411. doi: 10.1103/PhysRev.136.A405 (http://dx.doi.org/10.1103/PhysRev.136.A405).

[3] Bernal, J.D. (1964). "The Bakerian lecture, 1962: The structure of liquids". *Proc. R. Soc.* **280**: 299–322. doi: 10.1098/rspa.1964.0147 (http://dx.doi.org/10.1098/rspa.1964.0147).

[4] Schlick, T. (1996). "Pursuing Laplace's Vision on Modern Computers". in J. P. Mesirov, K. Schulten and D. W. Sumners. *Mathematical Applications to Biomolecular Structure and Dynamics, IMA Volumes in Mathematics and Its Applications*. **82**. New York: Springer-Verlag. pp. 218–247. ISBN 978-0387948386.

[5] de Laplace, P. S. (1820) (in French). *Oeuveres Completes de Laplace, Theorie Analytique des Probabilites*. Paris, France: Gauthier-Villars.

[6] Streett WB, Tildesley DJ, Saville G (1978). "Multiple time-step methods in molecular dynamics". *Mol Phys* **35** (3): 639-648. doi: 10.1080/00268977800100471 (http://dx.doi.org/10.1080/00268977800100471).

[7] Tuckerman ME, Berne BJ, Martyna GJ (1991). "Molecular dynamics algorithm for multiple time scales: systems with long range forces". *J Chem Phys* **94** (10): 6811-6815.

[8] Tuckerman ME, Berne BJ, Martyna GJ (1992). "Reversible multiple time scale molecular dynamics". *J Chem Phys* **97** (3): 1990-2001. doi: 10.1063/1.463137 (http://dx.doi.org/10.1063/1.463137).

[9] Sugita, Yuji; Yuko Okamoto (1999). "Replica-exchange molecular dynamics method for protein folding". *Chem Phys Letters* **314**: 141–151. doi: 10.1016/S0009-2614(99)01123-9 (http://dx.doi.org/10.1016/S0009-2614(99)01123-9).

[10] Brenner, D. W. (1990). "Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films". *Phys. Rev. B* **42** (15): 9458. doi: 10.1103/PhysRevB.42.9458 (http://dx.doi.org/10.1103/PhysRevB.42.9458).

[11] van Duin, A.; Siddharth Dasgupta, Francois Lorant and William A. Goddard III (2001). *J. Phys. Chem. A* **105**: 9398.

[12] Tersoff, J. (1989). ""Modeling solid-state chemistry: Interatomic potentials for multicomponent systems". *Phys. Rev. B* **39**: 5566. doi: 10.1103/PhysRevB.39.5566 (http://dx.doi.org/10.1103/PhysRevB.39.5566).

[13] Daw, M. S.; S. M. Foiles and M. I. Baskes (1993). "The embedded-atom method: a review of theory and applications". *Mat. Sci. And Engr. Rep.* **9**: 251. doi: 10.1016/0920-2307(93)90001-U (http://dx.doi.org/10.1016/0920-2307(93)90001-U).

[14] Cleri, F.; V. Rosato (1993). "Tight-binding potentials for transition metals and alloys". *Phys. Rev. B* **48**: 22. doi: 10.1103/PhysRevB.48.22 (http://dx.doi.org/10.1103/PhysRevB.48.22).

[15] Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD (2006). "A polarizable model of water for molecular dynamics simulations of biomolecules". *Chem Phys Lett* **418**: 245-249. doi: 10.1016/j.cplett.2005.10.135 (http://dx.doi.org/10.1016/j.cplett.2005.10.135).

[16] Patel, S.; MacKerell, Jr. AD; Brooks III, Charles L (2004). "CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model". *J Comput Chem* **25**: 1504–1514. doi: 10.1002/jcc.20077 (http://dx.doi.org/10.1002/jcc.20077).

[17] Billeter, SR; SP Webb, PK Agarwal, T Iordanov, S Hammes-Schiffer (2001). "Hydride Transfer in Liver Alcohol Dehydrogenase: Quantum Dynamics, Kinetic Isotope Effects, and Role of Enzyme Motion". *J Am Chem Soc* **123**: 11262–11272. doi: 10.1021/ja011384b (http://dx.doi.org/10.1021/ja011384b).

[18] Smith, A; CK Hall (2001). "Alpha-Helix Formation: Discontinuous Molecular Dynamics on an Intermediate-Resolution Protein Model". *Proteins* **44**: 344-360.

[19] Ding, F; JM Borreguero, SV Buldyrey, HE Stanley, NV Dokholyan (2003). "Mechanism for the alpha-helix to beta-hairpin transition". *J Am Chem Soc* **53**: 220–228. doi: 10.1002/prot.10468 (http://dx.doi.org/10.1002/prot.10468).

[20] Paci, E; M Vendruscolo, M Karplus (2002). "Validity of Go Models: Comparison with a Solvent-Shielded Empirical Energy Decomposition". *Biophys J* **83**: 3032–3038. doi: 10.1016/S0006-3495(02)75308-3 (http://dx.doi.org/10.1016/S0006-3495(02)75308-3).

[21] McCammon, J; JB Gelin, M Karplus (1977). "Dynamics of folded proteins". *Nature* **267**: 585–590. doi: 10.1038/267585a0 (http://dx.doi.org/10.1038/267585a0).

[22] Averback, R. S.; Diaz de la Rubia, T. (1998). "Displacement damage in irradiated metals and semiconductors". in H. Ehrenfest and F. Spaepen. *Solid State Physics*. **51**. New York: Academic Press. p. 281-402.

[23] R. Smith, ed (1997). *Atomic & ion collisions in solids and at surfaces: theory, simulation and applications*. Cambridge, UK: Cambridge University Press.

[24] Freddolino P, Arkhipov A, Larson SB, McPherson A, Schulten K. http://www.ks.uiuc.edu/Research/STMV/|"Molecular dynamics simulation of the Satellite Tobacco Mosaic Virus

(STMV)". *Theoretical and Computational Biophysics Group*. University of Illinois at Urbana Champaign. http://www.ks.uiuc.edu/Research/STMV/.
[25] The Folding@Home Project (http://folding.stanford.edu/) and recent papers (http://folding.stanford.edu/papers.html) published using trajectories from it. Vijay Pande Group. Stanford University
[26] http://www.cs.sandia.gov/~sjplimp/md.html
[27] http://cbbl.imim.es/Adun
[28] http://www.agilemolecule.com/Products.html
[29] http://cp2k.berlios.de/
[30] http://www.DEShawResearch.com/resources.html
[31] http://www.ccp5.ac.uk/DL_POLY/
[32] http://fireball-dft.org
[33] http://www.biowerkzeug.com/
[34] http://ccpforge.cse.rl.ac.uk/frs/?group_id=34
[35] http://www.univie.ac.at/newtonx/
[36] http://protomol.sourceforge.net/
[37] http://www.sphinxlib.de
[38] http://www.yasara.org
[39] http://www.chim.unifi.it/orac/
[40] http://esra.sourceforge.net/cgi-bin/index.cgi
[41] http://mw.concord.org/modeler/

General references

- M. P. Allen, D. J. Tildesley (1989) Computer simulation of liquids. Oxford University Press. ISBN 0-19-855645-4.
- J. A. McCammon, S. C. Harvey (1987) Dynamics of Proteins and Nucleic Acids. Cambridge University Press. ISBN 0521307503 (hardback).
- D. C. Rapaport (1996) The Art of Molecular Dynamics Simulation. ISBN 0-521-44561-2.
- Frenkel, Daan; Smit, Berend (2002) [2001]. *Understanding Molecular Simulation : from algorithms to applications*. San Diego, California: Academic Press. ISBN 0-12-267351-4.
- J. M. Haile (2001) Molecular Dynamics Simulation: Elementary Methods. ISBN 0-471-18439-X
- R. J. Sadus, *Molecular Simulation of Fluids: Theory, Algorithms and Object-Orientation*, 2002, ISBN 0-444-51082-6
- Oren M. Becker, Alexander D. Mackerell Jr, Benoît Roux, Masakatsu Watanabe (2001) Computational Biochemistry and Biophysics. Marcel Dekker. ISBN 0-8247-0455-X.
- Andrew Leach (2001) Molecular Modelling: Principles and Applications. (2nd Edition) Prentice Hall. ISBN 978-0582382107.
- Tamar Schlick (2002) Molecular Modeling and Simulation. Springer. ISBN 0-387-95404-X.
- William Graham Hoover (1991) Computational Statistical Mechanics, Elsevier, ISBN 0-444-88192-1.

External links

- The Blue Gene Project (http://researchweb.watson.ibm.com/bluegene/) (IBM)
- D. E. Shaw Research (http://deshawresearch.com/) (D. E. Shaw Research)
- Molecular Physics (http://www.tandf.co.uk/journals/titles/00268976.asp)
- Statistical mechanics of Nonequilibrium Liquids (http://www.phys.unsw.edu.au/~gary/book.html) Lecture Notes on non-equilibrium MD
- Introductory Lecture on Classical Molecular Dynamics (http://www.fz-juelich.de/nic-series/volume10/sutmann.pdf) by Dr. Godehard Sutmann, NIC, Forschungszentrum Jülich, Germany
- Introductory Lecture on Ab Initio Molecular Dynamics and Ab Initio Path Integrals (http://www.fz-juelich.de/nic-series/volume10/tuckerman2.pdf) by Mark E. Tuckerman, New York University, USA
- Introductory Lecture on Ab initio molecular dynamics: Theory and Implementation (http://www.fz-juelich.de/nic-series/Volume1/marx.pdf) by Dominik Marx, Ruhr-Universität Bochum and Jürg Hutter, Universität Zürich

CHARMM

Developer(s)	Martin Karplus, Accelrys
Initial release	1983
Stable release	c35b2 / 2008-12-28
Preview release	c36a2 / 2009-02-15
Written in	Ratfor
Operating system	Unix-like
Type	molecular dynamics
License	The CHARMM Development Project
Website	charmm.org [9]

CHARMM (**Chemistry at HARvard Macromolecular Mechanics**) is the name of a widely used set of force fields for molecular dynamics as well as the name for the molecular dynamics simulation and analysis package associated with them.[1] [2] The CHARMM Development Project involves a network of developers throughout the world working with Martin Karplus and his group at Harvard to develop and maintain the CHARMM program. Licenses for this software are available, for a fee, to people and groups working in academia.

The commercial version of CHARMM, called **CHARMm** (note the lowercase 'm'), is available from Accelrys.

CHARMM force fields

The CHARMM force fields for proteins include: united-atom (sometimes called "extended atom") CHARMM19[3] , all-atom CHARMM22[4] and its dihedral potential corrected variant CHARMM22/CMAP.[5] In the CHARMM22 protein force field, the atomic partial charges were derived from quantum chemical calculations of the interactions between model compounds and water. Furthermore, CHARMM22 is parametrized for the TIP3P explicit water model. Nevertheless, it is frequently used with implicit solvents. Recently, a special version of CHARMM22/CMAP was reparametrized for consistent use with implicit solvent GBSW.[6]

For DNA, RNA, and lipids, CHARMM27[7] is used. Some force fields may be combined, for example CHARMM22 and CHARMM27 for the simulation of protein-DNA binding. Additionally, parameters for NAD+, sugars, fluorinated compounds, etc. may be downloaded [8]. These force field version numbers refer to the CHARMM version where they first appeared, but may of course be used with subsequent versions of the CHARMM executable program. Likewise, these force fields may be used within other molecular dynamics programs that support them.

CHARMM also includes polarizable force fields using two approaches. One is based on the fluctuating charge (FQ) model, also known as Charge Equilibration (CHEQ). [9] [10] The other is based on the Drude shell or dispersion oscillator model. [11] [12]

CHARMM molecular dynamics program

The CHARMM program allows generation and analysis of a wide range of molecular simulations. The most basic kinds of simulation are minimization of a given structure and production runs of a molecular dynamics trajectory.

More advanced features include free energy perturbation (FEP), quasi-harmonic entropy estimation, correlation analysis and combined quantum, and molecular mechanics (QM/MM) methods.

CHARMM is one of the oldest programs for molecular dynamics. It has accumulated a huge number of features, some of which are duplicated under several keywords with slight variations. This is an inevitable result of the large number of outlooks and groups working on CHARMM throughout the world. The changelog file [13] as well as CHARMM's source code are good places to look for the names and affiliations of the main developers. The involvement and coordination by Charles L. Brooks III's group at the University of Michigan is very salient.

History of the program

Around 1969, there was considerable interest in developing potential energy functions for small molecules. CHARMM originated at Martin Karplus's group at Harvard. Karplus and his then graduate student Bruce Gelin decided the time was ripe to develop a program that would make it possible to take a given amino acid sequence and a set of coordinates (e.g., from the X-ray structure) and to use this information to calculate the energy of the system as a function of the atomic positions. Karplus has acknowledged the importance of major inputs in the development of the (still nameless) program, including

- Schneior Lifson's group at the Weizmann Institute, especially from Arieh Warshel who went to Harvard and brought his consistent force field (**CCF**) program with him;
- Harold Scheraga's group at Cornell University; and
- Awareness of Michael Levitt's pioneering energy calculations for proteins

In the 1980s, finally a paper appeared and CHARMM made its public début. Gelin's program had by then been considerably restructured. For the publication, Bob Bruccoleri came up with the name HARMM (HARvard Macromolecular Mechanics), but it didn't seem appropriate. So they added a C for Chemistry. Karplus said: "I sometimes wonder if Bruccoleri's original suggestion would have served as a useful warning to inexperienced scientists working with the program."[14] CHARMM has continued to grow and the latest release of the executable program was made in August 2008 as CHARMM35b1.

Running CHARMM Under Unix/Linux

The general syntax for using the program is:

```
charmm < filename.inp > filename.out
```

charmm

The actual name of the program (or script which runs the program) on the computer system being used.

filename.inp

A text file which contains the CHARMM commands. It starts by loading the molecular topologies (top) and force field (par). Then one loads the molecular structures' Cartesian coordinates (e.g. from PDB files). One can then modify the molecules (adding hydrogens, changing secondary structure). The calculation section can include energy minimization, dynamics production, and analysis tools such as motion and energy correlations.

`filename.out`

The log file for the CHARMM run, containing echoed commands, and various amounts of command output. The output print level may be increased or decreased in general, and procedures such as minimization and dynamics have printout frequency specifications. The values for temperature, energy pressure, etc. are output at that frequency.

CHARMM and Volunteer Computing

Docking@Home, hosted by University of Delaware, one of the projects which use a opensource platform for the distributed computing, BOINC, adopts CHARMM to analyze the atomic details of protein-ligand interactions in terms of Molecular Dynamics (MD) simulations and minimizations.

World Community Grid, sponsored by IBM, runs a project called The Clean Energy Project [15] which also uses CHARMM.

See also

- AMBER
- Force field implementation

References

[1] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations". *J Comp Chem* **4**: 187–217. doi: 10.1002/jcc.540040211 (http://dx.doi.org/10.1002/jcc.540040211).

[2] MacKerell, A.D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. (1998). "CHARMM: The Energy Function and Its Parameterization with an Overview of the Program". in Schleyer, P.v.R.; et al.. *The Encyclopedia of Computational Chemistry*. **1**. Chichester: John Wiley & Sons. pp. 271-277.

[3] Reiher, III WH (1985). "Theoretical studies of hydrogen bonding". *PhD Thesis at Harvard University*.

[4] MacKerell, Jr. AD, *et al.* (1998). "All-atom empirical potential for molecular modeling and dynamics studies of proteins". *J Phys Chem B* **102**: 3586–3616. doi: 10.1021/jp973084f (http://dx.doi.org/10.1021/jp973084f).

[5] MacKerell, Jr. AD, Feig M, Brooks, III CL (2004). "Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". *J Comput Chem* **25**: 1400–1415. doi: 10.1002/jcc.20065 (http://dx.doi.org/10.1002/jcc.20065).

[6] Brooks CL, Chen J, Im W (2006). "Balancing solvation and intramolecular interactions: toward a consistent generalized born force field (CMAP opt. for GBSW)". *J Am Chem Soc* **128**: 3728–3736. doi: 10.1021/ja057216r (http://dx.doi.org/10.1021/ja057216r).

[7] MacKerell, Jr. AD, Banavali N, Foloppe N (2001). "Development and current status of the CHARMM force field for nucleic acids". *Biopolymers* **56**: 257–265. doi: 10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W (http://dx.doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W).

[8] http://mackerell.umaryland.edu/CHARMM_ff_params.html

[9] Patel, S.; MacKerell, Jr. AD; Brooks III, Charles L (2004). "CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations". *J Comput Chem* **25**: 1–15. doi: 10.1002/jcc.10355 (http://dx.doi.org/10.1002/jcc.10355).

[10] Patel, S.; MacKerell, Jr. AD; Brooks III, Charles L (2004). "CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model". *J Comput Chem* **25**: 1504–1514. doi: 10.1002/jcc.20077 (http://dx.doi.org/10.1002/jcc.20077).
[11] Lamoureux G, Roux B. (2003). Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J Chem Phys* 119(6):3025-3039.
[12] Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD. (2006). A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem Phys Lett* 418:245-9.
[13] http://www.charmm.org/package/changelogs/c34log.shtml
[14] Karplus M (2006). "Spinach on the ceiling: a theoretical chemist's return to biology". *Annu Rev Biophys Biomol Struct* **35**: 1–47. doi: 10.1146/annurev.biophys.33.110502.133350 (http://dx.doi.org/10.1146/annurev.biophys.33.110502.133350).
[15] http://www.worldcommunitygrid.org/projects_showcase/cep1/viewCep1Main.do

External links

- Accelrys website (http://www.accelrys.com/)
- CHARMM website (http://www.charmm.org/) with documentation (http://www.charmm.org/html/documentation/chmdoc.html) and helpful discussion forums (http://165.112.184.13//ubbthreads/ubbthreads.php?Cat=)
- CHARMM tutorial (http://www.ch.embnet.org/MD_tutorial/)
- MacKerell (http://www.pharmacy.umaryland.edu/faculty/amackere/) website including a Package of force field parameters for CHARMM (http://mackerell.umaryland.edu/CHARMM_ff_params.html)
- C.Brooks website (http://www.scripps.edu/brooks/)
- CHARMM page at Harvard (http://yuri.harvard.edu/)
- Roux website (http://thallium.bsd.uchicago.edu/RouxLab/index.html)
- Bernard R. Brooks Group Website (http://www.lobos.nih.gov/cbs/index.php)
- VMD (http://www.ks.uiuc.edu/Research/vmd/) - visualization of CHARMM trajectories
- Sirius (http://sirius.sdsc.edu) - visualization of CHARMM trajectories
- Docking@Home (http://docking.cis.udel.edu/)

Statistical mechanics

Statistical mechanics (or **statistical thermodynamics**[1]) is the application of probability theory, which includes mathematical tools for dealing with large populations, to the field of mechanics, which is concerned with the motion of particles or objects when subjected to a force. It provides a framework for relating the microscopic properties of individual atoms and molecules to the macroscopic or bulk properties of materials that can be observed in everyday life, therefore explaining thermodynamics as a natural result of statistics and mechanics (classical and quantum) at the microscopic level.

It provides a molecular-level interpretation of thermodynamic quantities such as work, heat, free energy, and entropy, allowing the thermodynamic properties of bulk materials to be related to the spectroscopic data of individual molecules. This ability to make macroscopic predictions based on microscopic properties is the main advantage of statistical mechanics over classical thermodynamics. Both theories are governed by the second law of thermodynamics through the medium of entropy. However, entropy in thermodynamics can only be known empirically, whereas in statistical mechanics, it is a function of the distribution of the system on its micro-states.

Statistical thermodynamics was born in 1870 with the work of Austrian physicist Ludwig Boltzmann, much of which was collectively published in Boltzmann's 1896 *Lectures on Gas Theory*.[2] Boltzmann's original papers on the statistical interpretation of thermodynamics, the H-theorem, transport theory, thermal equilibrium, the equation of state of gases, and similar subjects, occupy about 2,000 pages in the proceedings of the Vienna Academy and other societies. The term "statistical thermodynamics" was proposed for use by the American thermodynamicist and physical chemist J. Willard Gibbs in 1902. According to Gibbs, the term "statistical", in the context of mechanics, i.e. statistical mechanics, was first used by the Scottish physicist James Clerk Maxwell in 1871.

Overview

The essential problem in statistical thermodynamics is to determine the distribution of a given amount of energy E over N identical systems.[3] The goal of statistical thermodynamics is to understand and to interpret the measurable macroscopic properties of materials in terms of the properties of their constituent particles and the interactions between them. This is done by connecting thermodynamic functions to quantum-mechanic equations. Two central quantities in statistical thermodynamics are the Boltzmann factor and the partition function.

Fundamentals

Central topics covered in statistical thermodynamics include:

- Microstates and configurations
- Boltzmann distribution law
- Partition function, Configuration integral or configurational partition function
- Thermodynamic equilibrium - thermal, mechanical, and chemical.
- Internal degrees of freedom - rotation, vibration, electronic excitation, etc.
- Heat capacity – Einstein solids, polyatomic gases, etc.
- Nernst heat theorem
- Fluctuations
- Gibbs paradox
- Degeneracy

Lastly, and most importantly, the formal definition of entropy of a thermodynamic system from a statistical perspective is called statistical entropy, and is defined as:

$$S = k_B \ln \Omega$$

where

k_B is Boltzmann's constant 1.38066×10^{-23} J K^{-1} and

$\Omega$is the number of microstates corresponding to the observed thermodynamic macrostate.

A common mistake is taking this formula as a hard general definition of entropy. This equation is valid only if each microstate is equally accessible (each microstate has an equal probability of occurring).

Boltzmann Distribution

If the system is large the Boltzmann distribution could be used (the Boltzmann distribution is an approximate result)

$$n_i \propto e^{-\frac{U_i}{k_B T}}.$$

This can now be used with $\rho_i = \frac{n_i}{N}$:

$$\rho_i = \frac{n_i}{N} = \frac{e^{-\frac{U_i}{k_B T}}}{\sum_{i=1}^{\text{all levels}} e^{-\frac{U_i}{k_B T}}}.$$

History

In 1738, Swiss physicist and mathematician Daniel Bernoulli published *Hydrodynamica* which laid the basis for the kinetic theory of gases. In this work, Bernoulli positioned the argument, still used to this day, that gases consist of great numbers of molecules moving in all directions, that their impact on a surface causes the gas pressure that we feel, and that what we experience as heat is simply the kinetic energy of their motion.

In 1859, after reading a paper on the diffusion of molecules by Rudolf Clausius, Scottish physicist James Clerk Maxwell formulated the Maxwell distribution of molecular velocities, which gave the proportion of molecules having a certain velocity in a specific range. This was the first-ever statistical law in physics.[4] Five years later, in 1864, Ludwig Boltzmann, a young student in Vienna, came across Maxwell's paper and was so inspired by it that he

spent much of his long and distinguished life developing the subject further.

Hence, the foundations of statistical thermodynamics were laid down in the late 1800s by those such as Maxwell, Ludwig Boltzmann, Max Planck, Rudolf Clausius, and Willard Gibbs who began to apply statistical and quantum atomic theory to ideal gas bodies. Predominantly, however, it was Maxwell and Boltzmann, working independently, who reached similar conclusions as to the statistical nature of gaseous bodies. Yet, one must consider Boltzmann to be the "father" of statistical thermodynamics with his 1875 derivation of the relationship between entropy S and multiplicity Ω, the number of microscopic arrangements (microstates) producing the same macroscopic state (macrostate) for a particular system.[5]

Fundamental postulate

The fundamental postulate in statistical mechanics (also known as the *equal a priori probability postulate*) is the following:

> *Given an isolated system in equilibrium, it is found with equal probability in each of its accessible microstates.*

This postulate is a fundamental assumption in statistical mechanics - it states that a system in equilibrium does not have any preference for any of its available microstates. Given Ω microstates at a particular energy, the probability of finding the system in a particular microstate is $p = 1/\Omega$.

This postulate is necessary because it allows one to conclude that for a system at equilibrium, the thermodynamic state (macrostate) which could result from the largest number of microstates is also the most probable macrostate of the system.

The postulate is justified in part, for classical systems, by Liouville's theorem (Hamiltonian), which shows that if the distribution of system points through accessible phase space is uniform at some time, it remains so at later times.

Similar justification for a discrete system is provided by the mechanism of detailed balance.

This allows for the definition of the *information function* (in the context of information theory):

$$I = -\sum_i \rho_i \ln \rho_i = \langle \ln \rho \rangle.$$

When all the probabilities (rhos) are equal, I is maximal, and we have minimal information about the system. When our information is maximal (i.e., one rho is equal to one and the rest to zero, such that we know what state the system is in), the function is minimal.

This "information function" is the same as the **reduced entropic function** in thermodynamics.

Statistical ensembles

Microcanonical ensemble

In microcanonical ensemble N, V and E are fixed. Since the second law of thermodynamics applies to isolated systems, the first case investigated will correspond to this case. The *Microcanonical ensemble* describes an isolated system.

The entropy of such a system can only increase, so that the maximum of its entropy corresponds to an equilibrium state for the system.

Because an isolated system keeps a constant energy, the total energy of the system does not fluctuate. Thus, the system can access only those of its micro-states that correspond to a given value *E* of the energy. The internal energy of the system is then strictly equal to its energy.

Let us call $\Omega(E)$ the number of micro-states corresponding to this value of the system's energy. The macroscopic state of maximal entropy for the system is the one in which all micro-states are equally likely to occur, with probability $1/\Omega(E)$, during the system's fluctuations.

$$S = -k_B \sum_{i=1}^{\Omega(E)} \left\{ \frac{1}{\Omega(E)} \ln \frac{1}{\Omega(E)} \right\} = k_B \ln \left(\Omega(E)\right)$$

where

S is the system entropy, and

k_B is Boltzmann's constant.

Canonical ensemble

In canonical ensemble N, V and T are fixed. Invoking the concept of the canonical ensemble, it is possible to derive the probability P_i that a macroscopic system in thermal equilibrium with its environment, will be in a given microstate with energy E_i according to the Boltzmann distribution:

$$P_i = \frac{e^{-\beta E_i}}{\sum_j^{j_{\max}} e^{-\beta E_j}}$$

where $\beta = \frac{1}{kT}$,

The temperature T arises from the fact that the system is in thermal equilibrium with its environment. The probabilities of the various microstates must add to one, and the normalization factor in the denominator is the canonical partition function:

$$Z = \sum_j^{j_{max}} e^{-\beta E_j}$$

where E_i is the energy of the i th microstate of the system. The partition function is a measure of the number of states accessible to the system at a given temperature. The article canonical ensemble contains a derivation of Boltzmann's factor and the form of the partition function from first principles.

To sum up, the probability of finding a system at temperature T in a particular state with energy E_i is

$$P_i = \frac{e^{-\beta E_i}}{Z}.$$

Thermodynamic Connection

The partition function can be used to find the expected (average) value of any microscopic property of the system, which can then be related to macroscopic variables. For instance, the expected value of the microscopic energy E is *interpreted* as the microscopic definition of the thermodynamic variable internal energy U, and can be obtained by taking the derivative of the partition function with respect to the temperature. Indeed,

$$\langle E\rangle = \frac{\sum_i E_i e^{-\beta E_i}}{Z} = -\frac{1}{Z}\frac{dZ}{d\beta}$$

implies, together with the interpretation of $\langle E\rangle$ as U, the following microscopic definition of internal energy:

$$U := -\frac{d\ln Z}{d\beta}.$$

The entropy can be calculated by (see Shannon entropy)

$$\frac{S}{k} = -\sum_i p_i \ln p_i = \sum_i \frac{e^{-\beta E_i}}{Z}(\beta E_i + \ln Z) = \ln Z + \beta U$$

which implies that

$$-\frac{\ln(Z)}{\beta} = U - TS = F$$

is the free energy of the system or in other words,

$$Z = e^{-\beta F}$$

Having microscopic expressions for the basic thermodynamic potentials U (internal energy), S (entropy) and F (free energy) is sufficient to derive expressions for other thermodynamic quantities. The basic strategy is as follows. There may be an intensive or extensive quantity that enters explicitly in the expression for the microscopic energy E_i, for instance magnetic field (intensive) or volume (extensive). Then, the conjugate thermodynamic variables are derivatives of the internal energy. The macroscopic magnetization (extensive) is the derivative of U with respect to the (intensive) magnetic field, and the pressure (intensive) is the derivative of U with respect to volume (extensive). The treatment in this section assumes no exchange of matter (i.e. fixed mass and fixed particle numbers). However, the volume of the system is variable which means the density is also variable.

This probability can be used to find the average value, which corresponds to the macroscopic value, of any property, J, that depends on the energetic state of the system by using the formula:

$$\langle J\rangle = \sum_i p_i J_i = \sum_i J_i \frac{e^{-\beta E_i}}{Z}$$

where $\langle J\rangle$ is the average value of property J. This equation can be applied to the internal energy, U:

$$U = \sum_i E_i \frac{e^{-\beta E_i}}{Z}$$

Subsequently, these equations can be combined with known thermodynamic relationships between U and V to arrive at an expression for pressure in terms of only temperature, volume and the partition function. Similar relationships in terms of the partition function can be derived for other thermodynamic properties as shown in the following table; see also

the detailed explanation in configuration integral [6].

Helmholtz free energy:	$F = -\frac{\ln Z}{\beta}$
Internal energy:	$U = -\left(\frac{\partial \ln Z}{\partial \beta}\right)_{N,V}$
Pressure:	$P = -\left(\frac{\partial F}{\partial V}\right)_{N,T} = \frac{1}{\beta}\left(\frac{\partial \ln Z}{\partial V}\right)_{N,T}$
Entropy:	$S = k(\ln Z + \beta U)$
Gibbs free energy:	$G = F + PV = -\frac{\ln Z}{\beta} + \frac{V}{\beta}\left(\frac{\partial \ln Z}{\partial V}\right)_{N,T}$
Enthalpy:	$H = U + PV$
Constant volume heat capacity:	$C_V = \left(\frac{\partial U}{\partial T}\right)_{N,V}$
Constant pressure heat capacity:	$C_P = \left(\frac{\partial H}{\partial T}\right)_{N,P}$
Chemical potential:	$\mu_i = -\frac{1}{\beta}\left(\frac{\partial \ln Z}{\partial N_i}\right)_{T,V,N}$

To clarify, this is not a grand canonical ensemble.

It is often useful to consider the energy of a given molecule to be distributed among a number of modes. For example, translational energy refers to that portion of energy associated with the motion of the center of mass of the molecule. Configurational energy refers to that portion of energy associated with the various attractive and repulsive forces between molecules in a system. The other modes are all considered to be internal to each molecule. They include rotational, vibrational, electronic and nuclear modes. If we assume that each mode is independent (a questionable assumption) the total energy can be expressed as the sum of each of the components:

$$E = E_t + E_c + E_n + E_e + E_r + E_v$$

Where the subscripts t, c, n, e, r, and v correspond to translational, configurational, nuclear, electronic, rotational and vibrational modes, respectively. The relationship in this equation can be substituted into the very first equation to give:

$$Z = \sum_i e^{-\beta(E_{ti}+E_{ci}+E_{ni}+E_{ei}+E_{ri}+E_{vi})}$$
$$= \sum_i e^{-\beta E_{ti}} e^{-\beta E_{ci}} e^{-\beta E_{ni}} e^{-\beta E_{ei}} e^{-\beta E_{ri}} e^{-\beta E_{vi}}$$

If we can assume all these modes are completely uncoupled and uncorrelated, so all these factors are in a probability sense completely independent, then

$$Z = Z_t Z_c Z_n Z_e Z_r Z_v$$

Thus a partition function can be defined for each mode. Simple expressions have been derived relating each of the various modes to various measurable molecular properties, such as the characteristic rotational or vibrational frequencies.

Expressions for the various molecular partition functions are shown in the following table.

Nuclear	$Z_n = 1 \qquad (T < 10^8 K)$

Electronic	$Z_e = W_0 e^{kTD_e + W_1 e^{-\theta_{e1}/T} + \cdots}$
Vibrational	$Z_v = \prod_j \frac{e^{-\theta_{vj}/2T}}{1 - e^{-\theta_{vj}/T}}$
Rotational (linear)	$Z_r = \frac{T}{\sigma}\theta_r$
Rotational (non-linear)	$Z_r = \frac{1}{\sigma}\sqrt{\frac{\pi T^3}{\theta_A \theta_B \theta_C}}$
Translational	$Z_t = \frac{(2\pi mkT)^{3/2}}{h^3}$
Configurational (ideal gas)	$Z_c = V$

These equations can be combined with those in the first table to determine the contribution of a particular energy mode to a thermodynamic property. For example the "rotational pressure" could be determined in this manner. The total pressure could be found by summing the pressure contributions from all of the individual modes, ie:

$$P = P_t + P_c + P_n + P_e + P_r + P_v$$

Grand canonical ensemble

In grand canonical ensemble V, T and chemical potential are fixed. If the system under study is an open system, (matter can be exchanged), *but* particle number is not conserved, we would have to introduce chemical potentials, μ_j, j = 1,...,n and replace the canonical partition function with the grand canonical partition function:

$$\Xi(V,T,\mu) = \sum_i \exp\left(\beta\left[\sum_{j=1}^{n}\mu_j N_{ij} - E_i\right]\right)$$

where N_{ij} is the number of j^{th} species particles in the i^{th} configuration. Sometimes, we also have other variables to add to the partition function, one corresponding to each conserved quantity. Most of them, however, can be safely interpreted as chemical potentials. In most condensed matter systems, things are nonrelativistic and mass is conserved. However, most condensed matter systems of interest also conserve particle number approximately (metastably) and the mass (nonrelativistically) is none other than the sum of the number of each type of particle times its mass. Mass is inversely related to density, which is the conjugate variable to pressure. For the rest of this article, we will ignore this complication and pretend chemical potentials don't matter. See grand canonical ensemble.

Let's rework everything using a grand canonical ensemble this time. The volume is left fixed and does not figure in at all in this treatment. As before, j is the index for those particles of species j and i is the index for microstate i:

$$U = \sum_i E_i \frac{\exp(-\beta(E_i - \sum_j \mu_j N_{ij}))}{\Xi}$$

$$N_j = \sum_i N_{ij} \frac{\exp(-\beta(E_i - \sum_j \mu_j N_{ij}))}{\Xi}$$

Grand potential:	$\Phi_G = -\frac{\ln \Xi}{\beta}$
Internal energy:	$U = -\left(\frac{\partial \ln \Xi}{\partial \beta}\right)_\mu + \sum_i \frac{\mu_i}{\beta}\left(\frac{\partial \ln \Xi}{\partial \mu_i}\right)_\beta$

Particle number:	$N_i = \frac{1}{\beta}\left(\frac{\partial \ln \Xi}{\partial \mu_i}\right)_\beta$
Entropy:	$S = k(\ln \Xi + \beta U - \beta \sum_i \mu_i N_i)$
Helmholtz free energy:	$F = \Phi_G + \sum_i \mu_i N_i = -\frac{\ln \Xi}{\beta} + \sum_i \frac{\mu_i}{\beta}\left(\frac{\partial \ln \Xi}{\partial \mu_i}\right)_\beta$

Equivalence between descriptions at the thermodynamic limit

All of the above descriptions differ in the way they allow the given system to fluctuate between its configurations.

In the micro-canonical ensemble, the system exchanges no energy with the outside world, and is therefore not subject to energy fluctuations; in the canonical ensemble, the system is free to exchange energy with the outside in the form of heat.

In the thermodynamic limit, which is the limit of large systems, fluctuations become negligible, so that all these descriptions converge to the same description. In other words, the macroscopic behavior of a system does not depend on the particular ensemble used for its description.

Given these considerations, the best ensemble to choose for the calculation of the properties of a macroscopic system is that ensemble which allows the result to be derived most easily.

Random walks

The study of long chain polymers has been a source of problems within the realms of statistical mechanics since about the 1950s. One of the reasons however that scientists were interested in their study is that the equations governing the behaviour of a polymer chain were independent of the chain chemistry. What is more, the governing equation turns out to be a random (diffusive) walk in space. Indeed, the Schrödinger equation is itself a diffusion equation in imaginary time, $t' = it$.

Random walks in time

The first example of a random walk is one in space, whereby a particle undergoes a random motion due to external forces in its surrounding medium. A typical example would be a pollen grain in a beaker of water. If one could somehow "dye" the path the pollen grain has taken, the path observed is defined as a random walk.

Consider a toy problem, of a train moving along a 1D track in the x-direction. Suppose that the train moves either a distance of + or - a fixed distance **b**, depending on whether a coin lands heads or tails when flipped. Lets start by considering the statistics of the steps the toy train takes (where S_i is the ith step taken):

$\langle S_i \rangle = 0$; due to *a priori* equal probabilities
$\langle S_i S_j \rangle = b^2 \delta_{ij}$.

The second quantity is known as the correlation function. The delta is the kronecker delta which tells us that if the indices *i* and *j* are different, then the result is 0, but if *i* = *j* then the kronecker delta is 1, so the correlation function returns a value of b^2. This makes sense, because if *i* = *j* then we are considering the same step. Rather trivially then it can be shown

that the average displacement of the train on the x-axis is 0;

$$x = \sum_{i=1}^{N} S_i$$
$$\langle x \rangle = \langle \sum_{i=1}^{N} S_i \rangle$$
$$\langle x \rangle = \sum_{i=1}^{N} \langle S_i \rangle.$$

As stated $\langle S_i \rangle$ is 0, so the sum of 0 is still 0. It can also be shown, using the same method demonstrated above, to calculate the root mean square value of problem. The result of this calculation is given below

$$x_{rms} = \sqrt{\langle x^2 \rangle} = b\sqrt{N}.$$

From the diffusion equation it can be shown that the distance a diffusing particle moves in a media is proportional to the root of the time the system has been diffusing for, where the proportionality constant is the root of the diffusion constant. The above relation, although cosmetically different reveals similar physics, where *N* is simply the number of steps moved (is loosely connected with time) and *b* is the characteristic step length. As a consequence we can consider diffusion as a random walk process.

Random walks in space

Random walks in space can be thought of as snapshots of the path taken by a random walker in time. One such example is the spatial configuration of long chain polymers.

There are two types of random walk in space: *self-avoiding random walks*, where the links of the polymer chain interact and do not overlap in space, and *pure random* walks, where the links of the polymer chain are non-interacting and links are free to lie on top of one another. The former type is most applicable to physical systems, but their solutions are harder to get at from first principles.

By considering a freely jointed, non-interacting polymer chain, the end-to-end vector is $\mathbf{R} = \sum_{i=1}^{N} \mathbf{r}_i$ where $\mathbf{r}_i$ is the vector position of the *i*-th link in the chain. As a result of the central limit theorem, if N >> 1 then we expect a Gaussian distribution for the end-to-end vector. We can also make statements of the statistics of the links themselves;
$\langle \mathbf{r}_i \rangle = 0$; by the isotropy of space
$\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = 3b^2\delta_{ij}$; all the links in the chain are uncorrelated with one another
Using the statistics of the individual links, it is easily shown that $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R} \cdot \mathbf{R} \rangle = 3Nb^2$. Notice this last result is the same as that found for random walks in time.

Assuming, as stated, that that distribution of end-to-end vectors for a very large number of identical polymer chains is gaussian, the probability distribution has the following form

$$P = \frac{1}{\left(\frac{2\pi Nb^2}{3}\right)^{3/2}} \exp \frac{-3\mathbf{R} \cdot \mathbf{R}}{2Nb^2}$$

What use is this to us? Recall that according to the principle of equally likely *a priori* probabilities, the number of microstates, Ω, at some physical value is directly proportional to the probability distribution at that physical value, *viz*;

$$\Omega(\mathbf{R}) = cP(\mathbf{R})$$

where c is an arbitrary proportionality constant. Given our distribution function, there is a maxima corresponding to $\mathbf{R} = 0$. Physically this amounts to there being more microstates which have an end-to-end vector of 0 than any other microstate. Now by considering

$S(\mathbf{R}) = k_B \ln \Omega(\mathbf{R})$

$\Delta S(\mathbf{R}) = S(\mathbf{R}) - S(0)$

$\Delta F = -T\Delta S(\mathbf{R})$

where *F* is the Helmholtz free energy it is trivial to show that

$$\Delta F = k_B T \frac{3R^2}{2Nb^2} = \frac{1}{2} K R^2 \quad ; K = \frac{3k_B T}{Nb^2}$$

A Hookian spring!

This result is known as the **Entropic Spring Result** and amounts to saying that upon stretching a polymer chain you are doing work on the system to drag it away from its (preferred) equilibrium state. An example of this is a common elastic band, composed of long chain (rubber) polymers. By stretching the elastic band you are doing work on the system and the band behaves like a conventional spring. What is particularly astonishing about this result however, is that the work done in stretching the polymer chain can be related entirely to the change in entropy of the system as a result of the stretching.

Classical thermodynamics vs. statistical thermodynamics

As an example, from a classical thermodynamics point of view one might ask what is it about a thermodynamic system of gas molecules, such as ammonia NH_3, that determines the free energy characteristic of that compound? Classical thermodynamics does not provide the answer. If, for example, we were given spectroscopic data, of this body of gas molecules, such as bond length, bond angle, bond rotation, and flexibility of the bonds in NH_3 we should see that the free energy could not be other than it is. To prove this true, we need to bridge the gap between the microscopic realm of atoms and molecules and the macroscopic realm of classical thermodynamics. From physics, statistical mechanics provides such a bridge by teaching us how to conceive of a thermodynamic *system* as an assembly of *units*. More specifically, it demonstrates how the thermodynamic parameters of a system, such as temperature and pressure, are interpretable in terms of the parameters descriptive of such constituent atoms and molecules.[7]

In a bounded system, the crucial characteristic of these microscopic units is that their energies are quantized. That is, where the energies accessible to a macroscopic system form a virtual continuum of possibilities, the energies open to any of its submicroscopic components are limited to a discontinuous set of alternatives associated with integral values of some quantum number.

See also

- Chemical thermodynamics
- Configuration entropy
- Dangerously irrelevant
- Paul Ehrenfest
- Equilibrium thermodynamics
- Fluctuation dissipation theorem
- Important Publications in Statistical Mechanics
- Ising Model
- Mean field theory
- Nanomechanics
- Non-equilibrium thermodynamics
- Quantum thermodynamics
- Statistical physics
- Thermochemistry
- Widom insertion method
- Monte Carlo method
- Molecular modelling

A Table of Statistical Mechanics Articles

	Maxwell Boltzmann	**Bose-Einstein**	**Fermi-Dirac**
Particle		Boson	Fermion
Statistics	Partition function Statistical properties Microcanonical ensemble | Canonical ensemble | Grand canonical ensemble		
Statistics	Maxwell-Boltzmann statistics Maxwell-Boltzmann distribution Boltzmann distribution Gibbs paradox	Bose-Einstein statistics	Fermi-Dirac statistics
Thomas-Fermi approximation	gas in a box gas in a harmonic trap		
Gas	Ideal gas	Bose gas Debye model Bose-Einstein condensate Planck's law of black body radiation	Fermi gas Fermion condensate
Chemical Equilibrium	Classical Chemical equilibrium		

Notes

[1] The terms "Statistical mechanics" and "statistical thermodynamics" are used interchangeably. "Statistical physics" is a broader term which includes statistical mechanics, but is sometimes also used as a synonym for statistical mechanics

[2] On history of fundamentals of statistical thermodynamics (http://www.worldscibooks.com/phy_etextbook/2012/2012_chap01.pdf) (section 1.2)

[3] Schrodinger, Erwin (1946). *Statistical Thermodynamics*. Dover Publications, Inc.. ISBN 0-486-66101-6. OCLC 20056858 (http://worldcat.org/oclc/20056858).

[4] Mahon, Basil (2003). *The Man Who Changed Everything – the Life of James Clerk Maxwell*. Hoboken, NJ: Wiley. ISBN 0-470-86171-1. OCLC 52358254 62045217 (http://worldcat.org/oclc/52358254+62045217).

[5] Perrot, Pierre (1998). *A to Z of Thermodynamics*. Oxford University Press. ISBN 0-19-856552-6. OCLC 123283342 38073404 (http://worldcat.org/oclc/123283342+38073404).

[6] http://clesm.mae.ufl.edu/wiki.pub/index.php/Configuration_integral_%28statistical_mechanics%29

[7] Nash, Leonard K. (1974). *Elements of Statistical Thermodynamics, 2nd Ed.*. Dover Publications, Inc.. ISBN 0-486-44978-5. OCLC 61513215 (http://worldcat.org/oclc/61513215).

References

- Chandler, David (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press. ISBN 0-19-504277-8. OCLC 13946448 (http://worldcat.org/oclc/13946448).
- Huang, Kerson (1990). *Statistical Mechanics*. Wiley, John & Sons. ISBN 0-471-81518-7. OCLC 15017884 (http://worldcat.org/oclc/15017884).
- Kittel, Charles; Herbert Kroemer (1980). *Thermal Physics* (2nd ed.). San Francisco: W.H. Freeman. ISBN 0716710889. OCLC 5171399 (http://worldcat.org/oclc/5171399).
- McQuarrie, Donald (2000). *Statistical Mechanics (2nd rev. Ed.)*. University Science Books. ISBN 1-891389-15-7. OCLC 43370175 (http://worldcat.org/oclc/43370175).
- Dill, Ken; Bromberg, Sarina (2003). *Molecular Driving Forces*. Garland Science. ISBN 0-8153-2051-5. OCLC 47915710 (http://worldcat.org/oclc/47915710).
- List of notable textbooks in statistical mechanics

Further reading

- Ben-Naim, Arieh (2007). *Statistical Thermodynamics Based on Information*. ISBN 978-981-270-707-9
- Boltzmann, Ludwig; and Dieter Flamm (2000). *Entropie und Wahrscheinlichkeit*. ISBN 978-3817132867
- Boltzmann, Ludwig (1896, 1898). *[Lectures on gas theory]*. New York: Dover. ISBN 0486684555. OCLC 31434905 (http://worldcat.org/oclc/31434905). translated by Stephen G. Brush (1964) Berkeley: University of California Press; (1995) New York: Dover ISBN 0-486-68455-5
- Gibbs, J. Willard (1981) [1902]. *Elementary principles in statistical dynamics*. Woodbridge, Connecticut: Ox Bow Press. ISBN 0-918024-20-X.
- Landau, Lev Davidovich; and Lifshitz, Evgeny Mikhailovich (1980) [1976]. *Statistical Physics*. **5** (3 ed.). Oxford: Pergamon Press. ISBN 0-7506-3372-7. Translated by J.B. Sykes and M.J. Kearsley
- Reichl, Linda E (1998) [1980]. *A modern course in statistical physics* (2 ed.). Chichester: Wiley. ISBN 0-471-59520-9.

External links

- Philosophy of Statistical Mechanics (http://plato.stanford.edu/entries/statphys-statmech/) article by Lawrence Sklar for the Stanford Encyclopedia of Philosophy.
- Sklogwiki - Thermodynamics, statistical mechanics, and the computer simulation of materials. (http://www.sklogwiki.org/) SklogWiki is particularly orientated towards liquids and soft condensed matter.
- Statistical Thermodynamics (http://history.hyperjeff.net/statmech.html) - Historical Timeline

Statistical field theory

A **statistical field theory** is any model in statistical mechanics where the degrees of freedom comprise a field or fields. In other words, the microstates of the system are expressed through field configurations. It is closely related to quantum field theory, which describes the quantum mechanics of fields, and shares with it many phenomena, such as renormalization. If the system involves polymers, it is also known as polymer field theory.

In fact, by performing a Wick rotation from Minkowski space to Euclidean space, many results of statistical field theory can be applied directly to its quantum equivalent. The correlation functions of a statistical field theory are called Schwinger functions, and their properties are described by the Osterwalder–Schrader axioms.

Statistical field theories are widely used to describe systems in polymer physics or biophysics, such as polymer films, nanostructured block copolymers[1] or polyelectrolytes[2] .

References

- *Statistical Field Theory* volumes I and II (Cambridge Monographs on Mathematical Physics) by Claude Itzykson, Jean-Michel Drouffe, Publisher: Cambridge University Press; (March 29, 1991) ISBN 0-521-40806-7 ISBN 0-521-40805-9
- *The $P(\varphi)_2$ Euclidean (quantum) field theory.* by Barry Simon. Princeton Univ Press (June 1974) ISBN 0-691-08144-1
- *Quantum Physics: A Functional Integral Point of View* by James Glimm, Jaffe. Springer; 2nd edition (May 1987) ISBN 0-387-96477-0

[1] Baeurle SA, Usami T, Gusev AA (2006). "A new multiscale modeling approach for the prediction of mechanical properties of polymer-based nanomaterials". *Polymer* **47**: 8604–8617. doi: 10.1016/j.polymer.2006.10.017 (http://dx.doi.org/10.1016/j.polymer.2006.10.017).

[2] Baeurle SA, Nogovitsin EA (2007). "Challenging scaling laws of flexible polyelectrolyte solutions with effective renormalization concepts". *Polymer* **48**: 4883–4899. doi: 10.1016/j.polymer.2007.05.080 (http://dx.doi.org/10.1016/j.polymer.2007.05.080).

External links

- Problems in Statistical Field Theory (http://www.gursey.gov.tr/~mgh/rg2006/problemsets.html)
- Particle and Polymer Field Theory Group (http://www-dick.chemie.uni-regensburg.de/group/stephan_baeurle/)

Computational chemistry

Computational chemistry is a branch of chemistry that uses computers to assist in solving chemical problems. It uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules and solids. While its results normally complement the information obtained by chemical experiments, it can in some cases predict hitherto unobserved chemical phenomena. It is widely used in the design of new drugs and materials.

Examples of such properties are structure (i.e. the expected positions of the constituent atoms), absolute and relative (interaction) energies, electronic charge distributions, dipoles and higher multipole moments, vibrational frequencies, reactivity or other spectroscopic quantities, and cross sections for collision with other particles.

The methods employed cover both static and dynamic situations. In all cases the computer time and other resources (such as memory and disk space) increase rapidly with the size of the system being studied. That system can be a single molecule, a group of molecules, or a solid. Computational chemistry methods range from highly accurate to very approximate; highly accurate methods are typically feasible only for small systems. Ab initio methods are based entirely on theory from first principles. Other (typically less accurate) methods are called empirical or semi-empirical because they employ experimental results, often from acceptable models of atoms or related molecules, to approximate some elements of the underlying theory.

Both ab initio and semi-empirical approaches involve approximations. These range from simplified forms of the first-principles equations that are easier or faster to solve, to approximations limiting the size of the system (for example, Periodic boundary conditions), to fundamental approximations to the underlying equations that are required to achieve any solution to them at all. For example, most ab initio calculations make the Born-Oppenheimer approximation, which greatly simplifies the underlying Schrödinger Equation by freezing the nuclei in place during the calculation. In principle, ab initio methods eventually converge to the exact solution of the underlying equations as the number of approximations is reduced. In practice, however, it is impossible to eliminate all approximations, and residual error inevitably remains. The goal of computational chemistry is to minimize this residual error while keeping the calculations tractable.

History

Building on the founding discoveries and theories in the history of quantum mechanics, the first theoretical calculations in chemistry were those of Walter Heitler and Fritz London in 1927. The books that were influential in the early development of computational quantum chemistry include: Linus Pauling and E. Bright Wilson’s 1935 *Introduction to Quantum Mechanics – with Applications to Chemistry*, Eyring, Walter and Kimball's 1944 *Quantum*

Chemistry, Heitler's 1945 *Elementary Wave Mechanics – with Applications to Quantum Chemistry*, and later Coulson's 1952 textbook *Valence*, each of which served as primary references for chemists in the decades to follow.

With the development of efficient computer technology in the 1940s, the solutions of elaborate wave equations for complex atomic systems began to be a realizable objective. In the early 1950s, the first semi-empirical atomic orbital calculations were carried out. Theoretical chemists became extensive users of the early digital computers. A very detailed account of such use in the United Kingdom is given by Smith and Sutcliffe.[1] The first ab initio Hartree-Fock calculations on diatomic molecules were carried out in 1956 at MIT, using a basis set of Slater orbitals. For diatomic molecules, a systematic study using a minimum basis set and the first calculation with a larger basis set were published by Ransil and Nesbet respectively in 1960.[2] The first polyatomic calculations using Gaussian orbitals were carried out in the late 1950s. The first configuration interaction calculations were carried out in Cambridge on the EDSAC computer in the 1950s using Gaussian orbitals by Boys and coworkers.[3] By 1971, when a bibliography of ab initio calculations was published,[4] the largest molecules included were naphthalene and azulene.[5] [6] Abstracts of many earlier developments in ab initio theory have been published by Schaefer.[7]

In 1964, Hückel method calculations (using a simple linear combination of atomic orbitals (LCAO) method for the determination of electron energies of molecular orbitals of π electrons in conjugated hydrocarbon systems) of molecules ranging in complexity from butadiene and benzene to ovalene, were generated on computers at Berkeley and Oxford.[8] These empirical methods were replaced in the 1960s by semi-empirical methods such as CNDO.[9]

In the early 1970s, efficient ab initio computer programs such as ATMOL, GAUSSIAN, IBMOL, and POLYAYTOM, began to be used to speed up ab initio calculations of molecular orbitals. Of these four programs, only GAUSSIAN, now massively expanded, is still in use, but many other programs are now in use. At the same time, the methods of molecular mechanics, such as MM2, were developed, primarily by Norman Allinger.[10]

One of the first mentions of the term "computational chemistry" can be found in the 1970 book *Computers and Their Role in the Physical Sciences* by Sidney Fernbach and Abraham Haskell Taub, where they state "It seems, therefore, that 'computational chemistry' can finally be more and more of a reality."[11] During the 1970s, widely different methods began to be seen as part of a new emerging discipline of *computational chemistry*.[12] The *Journal of Computational Chemistry* was first published in 1980.

Concepts

The term *theoretical chemistry* may be defined as a mathematical description of chemistry, whereas *computational chemistry* is usually used when a mathematical method is sufficiently well developed that it can be automated for implementation on a computer. Note that the words *exact* and *perfect* do not appear here, as very few aspects of chemistry can be computed exactly. However, almost every aspect of chemistry can be described in a qualitative or approximate quantitative computational scheme.

Molecules consist of nuclei and electrons, so the methods of quantum mechanics apply. Computational chemists often attempt to solve the non-relativistic Schrödinger equation, with relativistic corrections added, although some progress has been made in solving the fully relativistic Dirac equation. In principle, it is possible to solve the Schrödinger equation

in either its time-dependent or time-independent form, as appropriate for the problem in hand; in practice, this is not possible except for very small systems. Therefore, a great number of approximate methods strive to achieve the best trade-off between accuracy and computational cost. Accuracy can always be improved with greater computational cost. Significant errors can present themselves in *ab initio* models comprising many electrons, due to the computational expense of full relativistic-inclusive methods. This complicates the study of molecules interacting with high atomic mass unit atoms, such as transitional metals and their catalytic properties. Present algorithms in computational chemistry can routinely calculate the properties of molecules that contain up to about 40 electrons with sufficient accuracy. Errors for energies can be less than a few kJ/mol. For geometries, bond lengths can be predicted within a few picometres and bond angles within 0.5 degrees. The treatment of larger molecules that contain a few dozen electrons is computationally tractable by approximate methods such as density functional theory (DFT). There is some dispute within the field whether or not the latter methods are sufficient to describe complex chemical reactions, such as those in biochemistry. Large molecules can be studied by semi-empirical approximate methods. Even larger molecules are treated by classical mechanics methods that employ what are called molecular mechanics. In QM/MM methods, small portions of large complexes are treated quantum mechanically (QM), and the remainder is treated approximately (MM).

In theoretical chemistry, chemists, physicists and mathematicians develop algorithms and computer programs to predict atomic and molecular properties and reaction paths for chemical reactions. Computational chemists, in contrast, may simply apply existing computer programs and methodologies to specific chemical questions. There are two different aspects to computational chemistry:

- Computational studies can be carried out in order to find a starting point for a laboratory synthesis, or to assist in understanding experimental data, such as the position and source of spectroscopic peaks.
- Computational studies can be used to predict the possibility of so far entirely unknown molecules or to explore reaction mechanisms that are not readily studied by experimental means.

Thus, computational chemistry can assist the experimental chemist or it can challenge the experimental chemist to find entirely new chemical objects.

Several major areas may be distinguished within computational chemistry:

- The prediction of the molecular structure of molecules by the use of the simulation of forces, or more accurate quantum chemical methods, to find stationary points on the energy surface as the position of the nuclei is varied.
- Storing and searching for data on chemical entities (see chemical databases).
- Identifying correlations between chemical structures and properties (see QSPR and QSAR).
- Computational approaches to help in the efficient synthesis of compounds.
- Computational approaches to design molecules that interact in specific ways with other molecules (e.g. drug design and catalysis).

Methods

A single molecular formula can represent a number of molecular isomers. Each isomer is a local minimum on the energy surface (called the potential energy surface) created from the total energy (i.e., the electronic energy, plus the repulsion energy between the nuclei) as a function of the coordinates of all the nuclei. A stationary point is a geometry such that the derivative of the energy with respect to all displacements of the nuclei is zero. A local (energy) minimum is a stationary point where all such displacements lead to an increase in energy. The local minimum that is lowest is called the global minimum and corresponds to the most stable isomer. If there is one particular coordinate change that leads to a decrease in the total energy in both directions, the stationary point is a transition structure and the coordinate is the reaction coordinate. This process of determining stationary points is called geometry optimization.

The determination of molecular structure by geometry optimization became routine only after efficient methods for calculating the first derivatives of the energy with respect to all atomic coordinates became available. Evaluation of the related second derivatives allows the prediction of vibrational frequencies if harmonic motion is estimated. More importantly, it allows for the characterization of stationary points. The frequencies are related to the eigenvalues of the Hessian matrix, which contains second derivatives. If the eigenvalues are all positive, then the frequencies are all real and the stationary point is a local minimum. If one eigenvalue is negative (i.e., an imaginary frequency), then the stationary point is a transition structure. If more than one eigenvalue is negative, then the stationary point is a more complex one, and is usually of little interest. When one of these is found, it is necessary to move the search away from it if the experimenter is looking solely for local minima and transition structures.

The total energy is determined by approximate solutions of the time-dependent Schrödinger equation, usually with no relativistic terms included, and by making use of the Born-Oppenheimer approximation, which allows for the separation of electronic and nuclear motions, thereby simplifying the Schrödinger equation. This leads to the evaluation of the total energy as a sum of the electronic energy at fixed nuclei positions and the repulsion energy of the nuclei. A notable exception are certain approaches called direct quantum chemistry, which treat electrons and nuclei on a common footing. Density functional methods and semi-empirical methods are variants on the major theme. For very large systems, the relative total energies can be compared using molecular mechanics. The ways of determining the total energy to predict molecular structures are:

Ab initio methods

The programs used in computational chemistry are based on many different quantum-chemical methods that solve the molecular Schrödinger equation associated with the molecular Hamiltonian. Methods that do not include any empirical or semi-empirical parameters in their equations - being derived directly from theoretical principles, with no inclusion of experimental data - are called *ab initio* methods. This does not imply that the solution is an exact one; they are all approximate quantum mechanical calculations. It means that a particular approximation is rigorously defined on first principles (quantum theory) and then solved within an error margin that is qualitatively known beforehand. If numerical iterative methods have to be employed, the aim is to iterate until full machine accuracy is obtained (the best that is possible with a finite word length on the computer,

and within the mathematical and/or physical approximations made).

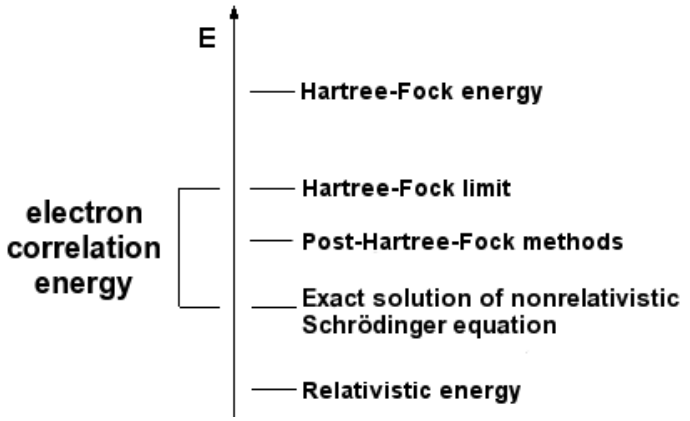


Diagram illustrating various *ab initio* electronic structure methods in terms of energy. Spacings are not to scale.

The simplest type of *ab initio* electronic structure calculation is the Hartree-Fock (HF) scheme, an extension of molecular orbital theory, in which the correlated electron-electron repulsion is not specifically taken into account; only its average effect is included in the calculation. As the basis set size is increased, the energy and wave function tend towards a limit called the Hartree-Fock limit. Many types of calculations (known as post-Hartree-Fock methods) begin with a Hartree-Fock calculation and subsequently correct for electron-electron repulsion, referred to also as electronic correlation. As these methods are pushed to the limit, they approach the exact solution of the non-relativistic Schrödinger equation. In order to obtain exact agreement with experiment, it is necessary to include relativistic and spin orbit terms, both of which are only really important for heavy atoms. In all of these approaches, in addition to the choice of method, it is necessary to choose a basis set. This is a set of functions, usually centered on the different atoms in the molecule, which are used to expand the molecular orbitals with the LCAO ansatz. Ab initio methods need to define a level of theory (the method) and a basis set.

The Hartree-Fock wave function is a single configuration or determinant. In some cases, particularly for bond breaking processes, this is quite inadequate, and several configurations need to be used. Here, the coefficients of the configurations and the coefficients of the basis functions are optimized together.

The total molecular energy can be evaluated as a function of the molecular geometry; in other words, the potential energy surface. Such a surface can be used for reaction dynamics. The stationary points of the surface lead to predictions of different isomers and the transition structures for conversion between isomers, but these can be determined without a full knowledge of the complete surface.

A particularly important objective, called computational thermochemistry, is to calculate thermochemical quantities such as the enthalpy of formation to chemical accuracy. Chemical accuracy is the accuracy required to make realistic chemical predictions and is generally considered to be 1 kcal/mol or 4 kJ/mol. To reach that accuracy in an economic way it is necessary to use a series of post-Hartree-Fock methods and combine the results. These methods are called quantum chemistry composite methods.

Density Functional methods

Density functional theory (DFT) methods are often considered to be *ab initio* methods for determining the molecular electronic structure, even though many of the most common functionals use parameters derived from empirical data, or from more complex calculations. In DFT, the total energy is expressed in terms of the total one-electron density rather than the wave function. In this type of calculation, there is an approximate Hamiltonian and an approximate expression for the total electron density. DFT methods can be very accurate for little computational cost. Some methods combine the density functional exchange functional with the Hartree-Fock exchange term and are known as hybrid functional methods.

Semi-empirical and empirical methods

Semi-empirical quantum chemistry methods are based on the Hartree-Fock formalism, but make many approximations and obtain some parameters from empirical data. They are very important in computational chemistry for treating large molecules where the full Hartree-Fock method without the approximations is too expensive. The use of empirical parameters appears to allow some inclusion of correlation effects into the methods.

Semi-empirical methods follow what are often called empirical methods, where the two-electron part of the Hamiltonian is not explicitly included. For π-electron systems, this was the Hückel method proposed by Erich Hückel, and for all valence electron systems, the Extended Hückel method proposed by Roald Hoffmann.

Molecular mechanics

In many cases, large molecular systems can be modeled successfully while avoiding quantum mechanical calculations entirely. Molecular mechanics simulations, for example, use a single classical expression for the energy of a compound, for instance the harmonic oscillator. All constants appearing in the equations must be obtained beforehand from experimental data or *ab initio* calculations.

The database of compounds used for parameterization, i.e., the resulting set of parameters and functions is called the force field, is crucial to the success of molecular mechanics calculations. A force field parameterized against a specific class of molecules, for instance proteins, would be expected to only have any relevance when describing other molecules of the same class.

These methods can be applied to proteins and other large biological molecules, and allow studies of the approach and interaction (docking) of potential drug molecules (eg. [13] and [14]).

Methods for solids

Computational chemical methods can be applied to solid state physics problems. The electronic structure of a crystal is in general described by a band structure, which defines the energies of electron orbitals for each point in the Brillouin zone. Ab initio and semi-empirical calculations yield orbital energies; therefore, they can be applied to band structure calculations. Since it is time-consuming to calculate the energy for a molecule, it is even more time-consuming to calculate them for the entire list of points in the Brillouin zone.

Chemical dynamics

Once the electronic and nuclear variables are separated (within the Born-Oppenheimer representation), in the time-dependent approach, the wave packet corresponding to the nuclear degrees of freedom is propagated via the time evolution operator (physics) associated to the time-dependent Schrödinger equation (for the full molecular Hamiltonian). In the complementary energy-dependent approach, the time-independent Schrödinger equation is solved using the scattering theory formalism. The potential representing the interatomic interaction is given by the potential energy surfaces. In general, the potential energy surfaces are coupled via the vibronic coupling terms.

The most popular methods for propagating the wave packet associated to the molecular geometry are

- the split operator technique,
- the Multi-Configuration Time-Dependent Hartree method (MCTDH),
- the semiclassical method.

Molecular dynamics (MD) examines (using Newton's laws of motion) the time-dependent behavior of systems, including vibrations or Brownian motion, using a classical mechanical description. MD combined with density functional theory leads to the Car-Parrinello method.

Interpreting molecular wave functions

The Atoms in Molecules model developed by Richard Bader was developed in order to effectively link the quantum mechanical picture of a molecule, as an electronic wavefunction, to chemically useful older models such as the theory of Lewis pairs and the valence bond model. Bader has demonstrated that these empirically useful models are connected with the topology of the quantum charge density. This method improves on the use of Mulliken population analysis.

Software packages

There are many self-sufficient software packages used by computational chemists. Some include many methods covering a wide range, while others concentrating on a very specific range or even a single method. Details of most of them can be found in:

- Quantum chemistry and solid state physics software supporting several methods.
- Molecular mechanics programs.
- Semi-empirical programs.
- Valence Bond programs.
- Biomolecular modelling programs: proteins, nucleic acid.

See also

- Mathematical chemistry
- Molecular modeling
- Molecular graphics
- Monte Carlo molecular modeling
- Quantum chemistry
- Basis set (chemistry)
- Molecular dynamics
- Bioinformatics
- Cheminformatics
- Computational Chemistry List
- Important publications in computational chemistry
- International Academy of Quantum Molecular Science
- Computational Science
- Statistical mechanics
- Molecule
- Force field in Chemistry
- Force field implementation

Cited References

[1] Smith, S. J.; Sutcliffe B. T., (1997). "The development of Computational Chemistry in the United Kingdom". *Reviews in Computational Chemistry* **70**: 271–316.

[2] Schaefer, Henry F. III (1972). *The electronic structure of atoms and molecules*. Reading, Massachusetss: Addison-Wesley Publishing Co.. pp. 146.

[3] Boys, S. F.; Cook G. B., Reeves C. M., Shavitt, I. (1956). "Automatic fundamental calculations of molecular structure". *Nature* **178** (2): 1207. doi: 10.1038/1781207a0 (http://dx.doi.org/10.1038/1781207a0).

[4] Richards, W. G.; Walker T. E. H and Hinkley R. K. (1971). *A bibliography of ab initio molecular wave functions*. Oxford: Clarendon Press.

[5] Preuss, H. (1968). International Journal of Quantum Chemistry 2: 651.

[6] Buenker, R. J.; Peyerimhoff S. D. (1969). Chemical Physics Letters 3: 37.

[7] Schaefer, Henry F. III (1984). *Quantum Chemistry*. Oxford: Clarendon Press.

[8] Streitwieser, A.; Brauman J. I. and Coulson C. A. (1965). *Supplementary Tables of Molecular Orbital Calculations*. Oxford: Pergamon Press.

[9] Pople, John A.; David L. Beveridge (1970). *Approximate Molecular Orbital Theory*. New York: McGraw Hill.

[10] Allinger, Norman (1977). "Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms". *Journal of the American Chemical Society* **99**: 8127–8134. doi: 10.1021/ja00467a001 (http://dx.doi.org/10.1021/ja00467a001).

[11] Fernbach, Sidney; Taub, Abraham Haskell (1970). *Computers and Their Role in the Physical Sciences*. Routledge. ISBN 0677140304.

[12] Reviews in Computational Chemistry vol 1, preface (http://www3.interscience.wiley.com/cgi-bin/bookhome/114034476)

[13] http://www.bio-balance.com/JMGM_article.pdf

[14] http://www.bio-balance.com/GPCR_Activation.pdf

Other references

- Christopher J. Cramer *Essentials of Computational Chemistry*, John Wiley & Sons (2002)
- T. Clark *A Handbook of Computational Chemistry*, Wiley, New York (1985)
- R. Dronskowski *Computational Chemistry of Solid State Materials*, Wiley-VCH (2005)
- F. Jensen *Introduction to Computational Chemistry*, John Wiley & Sons (1999)
- D. Rogers *Computational Chemistry Using the PC, 3rd Edition*, John Wiley & Sons (2003)
- Paul von Ragué Schleyer (Editor-in-Chief). *Encyclopedia of Computational Chemistry* (http://eu.wiley.com/WileyCDA/WileyTitle/productCd-047196588X.html). Wiley, **1998**. ISBN 0-471-96588-X.
- A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill (1982)
- D. Young *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons (2001)
- David Young's Introduction to Computational Chemistry (http://www.ccl.net/cca/documents/dyoung/topics-orig/compchem.html)
- K.I.Ramachandran, G Deepa and Krishnan Namboori. P.K. *Computational Chemistry and Molecular Modeling Principles and applications* (http://www.amrita.edu/cen/ccmm)Springer-Verlag GmbH ISBN 978-3-540-77302-3

External links

- NIST Computational Chemistry Comparison and Benchmark DataBase (http://srdata.nist.gov/cccbdb/) - Contains a database of thousands of computational and experimental results for hundreds of systems
- ACS Division of Computers in Chemistry (http://www.acscomp.org/) - ACS Computers in Chemistry Division
- Computational Chemistry Wiki (http://www.compchemwiki.org/index.php?title=Main_Page) - Wiki of computational chemistry results
- CSTB report (http://books.nap.edu/openbook.php?record_id=2206&page=R1) Mathematical Research in Materials Science: Opportunities and Perspectives - CSTB Report
- 3.320 Atomistic Computer Modeling of Materials (SMA 5107) (http://ocw.mit.edu/OcwWeb/Materials-Science-and-Engineering/3-320Spring-2005/CourseHome/) Free MIT Course
- Technology Roadmap for Computational Chemistry (http://www.chemicalvision2020.org/pdfs/compchem.pdf)
- APPLICATIONS OF MOLECULAR AND MATERIALS MODELING. (http://www.wtec.org/molmodel/mm_final.pdf)
- Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology CSTB Report (http://books.nap.edu/openbook.php?record_id=9591&page=1)

Mathematical chemistry

Mathematical chemistry is the area of research engaged in the novel and nontrivial applications of mathematics to chemistry; it concerns itself principally with the mathematical modeling of chemical phenomena.[1] Mathematical chemistry has also sometimes been called **computer chemistry**, but should not be confused with computational chemistry.

Major areas of research in mathematical chemistry include chemical graph theory, which deals with topics such as the mathematical study of isomerism and the development of topological descriptors or indices which find application in quantitative structure-property relationships; chemical aspects of group theory, which finds applications in stereochemistry and quantum chemistry; and topological aspects of chemistry.

The history of the approach may be traced back into 18th century. Georg Helm published a treatise titled "The Principles of Mathematical Chemistry: The Energetics of Chemical Phenomena" in 1894[2] . Some of the more contemporary periodical publications specializing in the field are MATCH Communications in Mathematical and in Computer Chemistry, first published in 1975, and the Journal of Mathematical Chemistry, first published in 1987.

The basic models for mathematical chemistry are molecular graph and topological index.

See also

- Cheminformatics
- Computational chemistry
- Combinatorial chemistry
- Molecular modeling

Bibliography

- Chemical Applications of Topology and Graph Theory, ed. by R. B. King, Elsevier, 1983.
- Mathematical Concepts in Organic Chemistry, by I. Gutman, O. E. Polansky, Springer-Verlag, Berlin, 1986.
- Mathematical Chemistry Series, ed. by D. Bonchev, D. H. Rouvray, Gordon and Breach Science Publisher, Amsterdam, 2000.

Notes

[1] A review (http://links.jstor.org/sici?sici=0036-1445(198806)30:2<348:MCIOC>2.0.CO;2-N&size=SMALL&origin=JSTOR-reducePage) of the book by Ivan Gutman, Oskar E. Polansky, "Mathematical Concepts in Organic Chemistry" in *SIAM Review* Vol. 30, No. 2 (1988), pp. 348-350

[2] Helm, Georg. The Principles of Mathematical Chemistry: The Energetics of Chemical Phenomena. translated by J. Livingston R. Morgan. New York: John Wiley & Sons, 1897. (http://books.google.com/books?hl=en&id=cyUuAAAAYAAJ&dq=helm+mathematical+chemistry&printsec=frontcover&source=web&ots=_0vgt-Wots&sig=w4EkV9aqaFd2jb6dXOOs3XObc9o&sa=X&oi=book_result&resnum=1&ct=result#PPP9,M1)

References

- N. Trinajstić, I. Gutman, Mathematical Chemistry, Croatica Chemica Acta, 75(2002), pp. 329-356.
- A. T. Balaban, Reflections about Mathematical Chemistry, Foundations of Chemistry, 7(2005), pp. 289-306.

External links

- Journal of Mathematical Chemistry (http://www.springerlink.com/content/101749/)
- MATCH Communications in Mathematical and in Computer Chemistry (http://www.pmf.kg.ac.yu/match/)

Monte Carlo method

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results. Monte Carlo methods are often used when simulating physical and mathematical systems. Because of their reliance on repeated computation and random or pseudo-random numbers, Monte Carlo methods are most suited to calculation by a computer. Monte Carlo methods tend to be used when it is unfeasible or impossible to compute an exact result with a deterministic algorithm.[1]

Monte Carlo simulation methods are especially useful in studying systems with a large number of coupled degrees of freedom, such as fluids, disordered materials, strongly coupled solids, and cellular structures (see cellular Potts model). More broadly, Monte Carlo methods are useful for modeling phenomena with significant uncertainty in inputs, such as the calculation of risk in business. These methods are also widely used in mathematics: a classic use is for the evaluation of definite integrals, particularly multidimensional integrals with complicated boundary conditions. It is a widely successful method in risk analysis when compared to alternative methods or human intuition. When Monte Carlo simulations have been applied in space exploration and oil exploration, actual observations of failures, cost overruns and schedule overruns are routinely better predicted by the simulations than by human intuition or alternative "soft" methods.[2]

The term "Monte Carlo method" was coined in the 1940s by physicists working on nuclear weapon projects in the Los Alamos National Laboratory.[3]

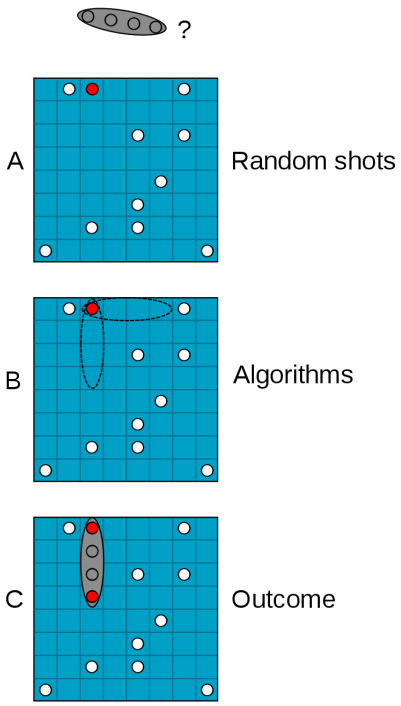
Overview

There is no single Monte Carlo method; instead, the term describes a large and widely-used class of approaches. However, these approaches tend to follow a particular pattern:

1. Define a domain of possible inputs.
2. Generate inputs randomly from the domain.
3. Perform a deterministic computation using the inputs.
4. Aggregate the results of the individual computations into the final result.

For example, the value of π can be approximated using a Monte Carlo method:

1. Draw a square on the ground, then inscribe a circle within it. From plain geometry, the ratio of the area of an inscribed circle to that of the surrounding square is π/4.
2. Uniformly scatter some objects of uniform size throughout the square. For example, grains of rice or sand.
3. Since the two areas are in the ratio π/4, the objects should fall in the areas in approximately the same ratio. Thus, counting the number of objects in the circle and dividing by the total number of objects in the square will yield an approximation for π/4. Multiplying the result by 4 will then yield an approximation for π itself.



The Monte Carlo method can be illustrated as a game of battleship. First a player makes some random shots. Next the player applies algorithms (i.e. a battleship is four dots in the vertical or horizontal direction). Finally based on the outcome of the random sampling and the algorithm the player can determine the likely locations of the other player's ships.

Notice how the π approximation follows the general pattern of Monte Carlo algorithms. First, we define a domain of inputs: in this case, it's the square which circumscribes our circle. Next, we generate inputs randomly (scatter individual grains within the square), then perform a computation on each input (test whether it falls within the circle). At the end, we aggregate the results into our final result, the approximation of π. Note, also, two other common properties of Monte Carlo methods: the computation's reliance on good random numbers, and its slow convergence to a better approximation as more data points are sampled. If grains are purposefully dropped into only, for example, the center of the circle, they will not be uniformly distributed, and so our approximation will be poor. An approximation will also be poor if only a few grains are randomly dropped into the whole square. Thus, the approximation of π will become more accurate both as the grains are dropped more uniformly and as more are dropped.

History

The name "Monte Carlo" was popularized by physics researchers Stanislaw Ulam, Enrico Fermi, John von Neumann, and Nicholas Metropolis, among others; the name is a reference to the Monte Carlo Casino in Monaco where Ulam's uncle would borrow money to gamble.[4] The use of randomness and the repetitive nature of the process are analogous to the activities conducted at a casino.

Random methods of computation and experimentation (generally considered forms of stochastic simulation) can be arguably traced back to the earliest pioneers of probability theory (see, e.g., Buffon's needle, and the work on small samples by William Sealy Gosset), but are more specifically traced to the pre-electronic computing era. The general difference usually described about a Monte Carlo form of simulation is that it systematically "inverts" the typical mode of simulation, treating deterministic problems by *first* finding a probabilistic analog (see Simulated annealing). Previous methods of simulation and statistical sampling generally did the opposite: using simulation to test a previously understood deterministic problem. Though examples of an "inverted" approach do exist historically, they were not considered a general method until the popularity of the Monte Carlo method spread.

Perhaps the most famous early use was by Enrico Fermi in 1930, when he used a random method to calculate the properties of the newly-discovered neutron. Monte Carlo methods were central to the simulations required for the Manhattan Project, though were severely limited by the computational tools at the time. Therefore, it was only after electronic computers were first built (from 1945 on) that Monte Carlo methods began to be studied in depth. In the 1950s they were used at Los Alamos for early work relating to the development of the hydrogen bomb, and became popularized in the fields of physics, physical chemistry, and operations research. The Rand Corporation and the U.S. Air Force were two of the major organizations responsible for funding and disseminating information on Monte Carlo methods during this time, and they began to find a wide application in many different fields.

Uses of Monte Carlo methods require large amounts of random numbers, and it was their use that spurred the development of pseudorandom number generators, which were far quicker to use than the tables of random numbers which had been previously used for statistical sampling.

Applications

As mentioned, Monte Carlo simulation methods are especially useful for modeling phenomena with significant uncertainty in inputs and in studying systems with a large number of coupled degrees of freedom. Specific areas of application include:

Physical sciences

Monte Carlo methods are very important in computational physics, physical chemistry, and related applied fields, and have diverse applications from complicated quantum chromodynamics calculations to designing heat shields and aerodynamic forms. The Monte Carlo method is widely used in statistical physics, in particular, Monte Carlo molecular modeling as an alternative for computational molecular dynamics; see Monte Carlo method in statistical physics. In experimental particle physics, these methods are used for

designing detectors, understanding their behavior and comparing experimental data to theory.

Monte Carlo methods are also used in the ensemble models that form the basis of modern weather forecasting operations.

Design and visuals

Monte Carlo methods have also proven efficient in solving coupled integral differential equations of radiation fields and energy transport, and thus these methods have been used in global illumination computations which produce photorealistic images of virtual 3D models, with applications in video games, architecture, design, computer generated films, special effects in cinema.

Finance and business

Monte Carlo methods in finance are often used to calculate the value of companies, to evaluate investments in projects at corporate level or to evaluate financial derivatives. The Monte Carlo method is intended for financial analysts who want to construct stochastic or probabilistic financial models as opposed to the traditional static and deterministic models. For its use in the insurance industry, see stochastic modelling.

Telecommunications

When planning a wireless network, design must be proved to work for a wide variety of scenarios that depend mainly on the number of users, their locations and the services they want to use. Monte Carlo methods are typically used to generate these users and their states. The network performance is then evaluated and, if results are not satisfactory, the network design goes through an optimization process.

Games

Monte Carlo methods have recently been applied in game playing related artificial intelligence theory. Most notably the game of Go has seen remarkably successful Monte Carlo algorithm based computer players. One of the main problems that this approach has in game playing is that it sometimes misses an isolated, very good move. These approaches are often strong strategically but weak tactically, as tactical decisions tend to rely on a small number of crucial moves which are easily missed by the randomly searching Monte Carlo algorithm.

Monte Carlo simulation versus “what if” scenarios

The opposite of Monte Carlo simulation might be considered deterministic modeling using single-point estimates. Each uncertain variable within a model is assigned a “best guess” estimate. Various combinations of each input variable are manually chosen (such as best case, worst case, and most likely case), and the results recorded for each so-called “what if” scenario. [5]

By contrast, Monte Carlo simulation considers random sampling of probability distribution functions as model inputs to produce hundreds or thousands of possible outcomes instead of a few discrete scenarios. The results provide probabilities of different outcomes occurring. [6] For example, a comparison of a spreadsheet cost construction model run using traditional “what if” scenarios, and then run again with Monte Carlo simulation and

Triangular probability distributions shows that the Monte Carlo analysis has a narrower range than the “what if” analysis. This is because the “what if” analysis gives equal weight to all scenarios.[7]

For an application, see quantifying uncertainty under corporate finance.

Use in mathematics

In general, Monte Carlo methods are used in mathematics to solve various problems by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties. The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically. The most common application of the Monte Carlo method is Monte Carlo integration.

Integration

Deterministic methods of numerical integration operate by taking a number of evenly spaced samples from a function. In general, this works very well for functions of one variable. However, for functions of vectors, deterministic quadrature methods can be very inefficient. To numerically integrate a function of a two-dimensional vector, equally spaced grid points over a two-dimensional surface are required. For instance a 10x10 grid requires 100 points. If the vector has 100 dimensions, the same spacing on the grid would require 10^{100} points—far too many to be computed. 100 dimensions is by no means unreasonable, since in many physical problems, a "dimension" is equivalent to a degree of freedom. (See Curse of dimensionality.)

Monte Carlo methods provide a way out of this exponential time-increase. As long as the function in question is reasonably well-behaved, it can be estimated by randomly selecting points in 100-dimensional space, and taking some kind of average of the function values at these points. By the law of large numbers, this method will display $1/\sqrt{N}$ convergence—i.e. quadrupling the number of sampled points will halve the error, regardless of the number of dimensions.

A refinement of this method is to somehow make the points random, but more likely to come from regions of high contribution to the integral than from regions of low contribution. In other words, the points should be drawn from a distribution similar in form to the integrand. Understandably, doing this precisely is just as difficult as solving the integral in the first place, but there are approximate methods available: from simply making up an integrable function thought to be similar, to one of the adaptive routines discussed in the topics listed below.

A similar approach involves using low-discrepancy sequences instead—the quasi-Monte Carlo method. Quasi-Monte Carlo methods can often be more efficient at numerical integration because the sequence "fills" the area better in a sense and samples more of the most important points that can make the simulation converge to the desired solution more quickly.

Integration methods

- Direct sampling methods
  - Importance sampling
  - Stratified sampling
  - Recursive stratified sampling
  - VEGAS algorithm
- Random walk Monte Carlo including Markov chains
  - Metropolis-Hastings algorithm
- Gibbs sampling

Optimization

Another powerful and very popular application for random numbers in numerical simulation is in numerical optimization. These problems use functions of some often large-dimensional vector that are to be minimized (or maximized). Many problems can be phrased in this way: for example a computer chess program could be seen as trying to find the optimal set of, say, 10 moves which produces the best evaluation function at the end. The traveling salesman problem is another optimization problem. There are also applications to engineering design, such as multidisciplinary design optimization.

Most Monte Carlo optimization methods are based on random walks. Essentially, the program will move around a marker in multi-dimensional space, tending to move in directions which lead to a lower function, but sometimes moving against the gradient.

Optimization methods

- Evolution strategy
- Genetic algorithms
- Parallel tempering
- Simulated annealing
- Stochastic optimization
- Stochastic tunneling

Inverse problems

Probabilistic formulation of inverse problems leads to the definition of a probability distribution in the model space. This probability distribution combines *a priori* information with new information obtained by measuring some observable parameters (data). As, in the general case, the theory linking data with model parameters is nonlinear, the a posteriori probability in the model space may not be easy to describe (it may be multimodal, some moments may not be defined, etc.).

When analyzing an inverse problem, obtaining a maximum likelihood model is usually not sufficient, as we normally also wish to have information on the resolution power of the data. In the general case we may have a large number of model parameters, and an inspection of the marginal probability densities of interest may be impractical, or even useless. But it is possible to pseudorandomly generate a large collection of models according to the posterior probability distribution and to analyze and display the models in such a way that information on the relative likelihoods of model properties is conveyed to the spectator. This can be accomplished by means of an efficient Monte Carlo method, even in cases where no explicit formula for the a priori distribution is available.

The best-known importance sampling method, the Metropolis algorithm, can be generalized, and this gives a method that allows analysis of (possibly highly nonlinear) inverse problems with complex a priori information and data with an arbitrary noise distribution. For details, see Mosegaard and Tarantola (1995),[8] or Tarantola (2005).[9]

Computational mathematics

Monte Carlo methods are useful in many areas of computational mathematics, where a *lucky choice* can find the correct result. A classic example is Rabin's algorithm for primality testing: for any n which is not prime, a random x has at least a 75% chance of proving that n is not prime. Hence, if n is not prime, but x says that it might be, we have observed at most a 1-in-4 event. If 10 different random x say that "n is probably prime" when it is not, we have observed a one-in-a-million event. In general a Monte Carlo algorithm of this kind produces one correct answer with a guarantee ***n* is composite, and *x* proves it so**, but another one without, but with a guarantee of not getting this answer when it is wrong **too often** — in this case at most 25% of the time. See also Las Vegas algorithm for a related, but different, idea.

Monte Carlo and random numbers

Interestingly, Monte Carlo simulation methods do not always require truly random numbers to be useful — while for some applications, such as primality testing, unpredictability is vital (see Davenport (1995)).[10] Many of the most useful techniques use deterministic, pseudo-random sequences, making it easy to test and re-run simulations. The only quality usually necessary to make good simulations is for the pseudo-random sequence to appear "random enough" in a certain sense.

What this means depends on the application, but typically they should pass a series of statistical tests. Testing that the numbers are uniformly distributed or follow another desired distribution when a large enough number of elements of the sequence are considered is one of the simplest, and most common ones.

See also

General

- Auxiliary field Monte Carlo
- Bootstrapping (statistics)
- Demon algorithm
- Evolutionary Computation
- Las Vegas algorithm
- Markov chain
- Molecular dynamics
- Monte Carlo option model
- Monte Carlo integration
- Quasi-Monte Carlo method
- Random number generator
- Randomness
- Resampling (statistics)

Application areas

- Graphics, particularly for ray tracing; a version of the Metropolis-Hastings algorithm is also used for ray tracing where it is known as Metropolis light transport
- Modeling light transport in biological tissue
- Monte Carlo methods in finance
- Reliability engineering
- In simulated annealing for protein structure prediction
- In semiconductor device research, to model the transport of current carriers
- Environmental science, dealing with contaminant behavior
- Search And Rescue and Counter-Pollution. Models used to predict the drift of a life raft or movement of an oil slick at sea.
- In probabilistic design for simulating and understanding the effects of variability
- In physical chemistry, particularly for simulations involving atomic clusters
- In biomolecular simulations
- In polymer physics
  - Bond fluctuation model
- In computer science
  - Las Vegas algorithm
  - LURCH
  - Computer go
  - General Game Playing
- Modeling the movement of impurity atoms (or ions) in plasmas in existing and tokamaks (e.g.: DIVIMP).
- Nuclear and particle physics codes using the Monte Carlo method:
  - GEANT — CERN's simulation of high energy particles interacting with a detector.
  - CompHEP, PYTHIA — Monte-Carlo generators of particle collisions
  - MCNP(X) - LANL's radiation transport codes
  - MCU: universal computer code for simulation of particle transport (neutrons, photons, electrons) in three-dimensional systems by means of the Monte Carlo method
  - EGS — Stanford's simulation code for coupled transport of electrons and photons
  - PEREGRINE: LLNL's Monte Carlo tool for radiation therapy dose calculations
  - BEAMnrc — Monte Carlo code system for modeling radiotherapy sources (LINAC's)
  - PENELOPE — Monte Carlo for coupled transport of photons and electrons, with applications in radiotherapy
  - MONK — Serco Assurance's code for the calculation of k-effective of nuclear systems
- Modelling of foam and cellular structures
- Modeling of tissue morphogenesis
- Computation of holograms
- Phylogenetic analysis, i.e. Bayesian inference, Markov chain Monte Carlo

Other methods employing Monte Carlo

- Assorted random models, e.g. self-organised criticality
- Direct simulation Monte Carlo
- Dynamic Monte Carlo method
- Kinetic Monte Carlo
- Quantum Monte Carlo
- Quasi-Monte Carlo method using low-discrepancy sequences and self avoiding walks
- Semiconductor charge transport and the like
- Electron microscopy beam-sample interactions
- Stochastic optimization
- Cellular Potts model
- Markov chain Monte Carlo
- Cross-entropy method
- Applied information economics
- Monte Carlo localization

Notes

[1] Douglas Hubbard "How to Measure Anything: Finding the Value of Intangibles in Business" pg. 46, John Wiley & Sons, 2007

[2] Douglas Hubbard "The Failure of Risk Management: Why It's Broken and How to Fix It", John Wiley & Sons, 2009

[3] Nicholas Metropolis (1987), "http://library.lanl.gov/la-pubs/00326866.pdf|The beginning of the Monte Carlo method", *Los Alamos Science* (1987 Special Issue dedicated to Stanislaw Ulam): 125–130, http://library.lanl.gov/la-pubs/00326866.pdf

[4] Douglas Hubbard "How to Measure Anything: Finding the Value of Intangibles in Business" pg. 46, John Wiley & Sons, 2007

[5] David Vose: “Risk Analysis, A Quantitative Guide,” Second Edition, p. 13, John Wiley & Sons, 2000.

[6] Ibid, p. 16

[7] Ibid, p. 17, showing graph

[8] http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/Papers_PDF/MonteCarlo_latex.pdf

[9] http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/SIAM/index.html

[10] Davenport, J. H.. http://doi.acm.org/10.1145/143242.143290|"Primality testing revisited". doi: http://doi.acm.org/10.1145/143242.143290 (http://dx.doi.org/http://doi.acm.org/10.1145/143242.143290). http://doi.acm.org/10.1145/143242.143290. Retrieved on 2007-08-19.

References

- Metropolis, N.; Ulam, S. (1949). "The Monte Carlo Method". *Journal of the American Statistical Association* **44** (247): 335–341. doi: 10.2307/2280232 (http://dx.doi.org/10.2307/2280232).
- Metropolis, Nicholas; Rosenbluth, Arianna W.; Rosenbluth, Marshall N.; Teller, Augusta H.; Teller, Edward (1953). "Equation of State Calculations by Fast Computing Machines". *Journal of Chemical Physics* **21** (6): 1087. doi: 10.1063/1.1699114 (http://dx.doi.org/10.1063/1.1699114).
- Hammersley, J. M.; Handscomb, D. C. (1975). *Monte Carlo Methods*. London: Methuen. ISBN 0416523404.
- Kahneman, D.; Tversky, A. (1982). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Gould, Harvey; Tobochnik, Jan (1988). *An Introduction to Computer Simulation Methods, Part 2, Applications to Physical Systems*. Reading: Addison-Wesley. ISBN 020116504X.

- Binder, Kurt (1995). *The Monte Carlo Method in Condensed Matter Physics*. New York: Springer. ISBN 0387543694.
- Berg, Bernd A. (2004). *Markov Chain Monte Carlo Simulations and Their Statistical Analysis (With Web-Based Fortran Code)*. Hackensack, NJ: World Scientific. ISBN 9812389350.
- Caflisch, R. E. (1998). *Monte Carlo and quasi-Monte Carlo methods*. Acta Numerica. **7**. Cambridge University Press. pp. 1–49.
- Doucet, Arnaud; Freitas, Nando de; Gordon, Neil (2001). *Sequential Monte Carlo methods in practice*. New York: Springer. ISBN 0387951466.
- Fishman, G. S. (1995). *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer. ISBN 038794527X.
- MacKeown, P. Kevin (1997). *Stochastic Simulation in Physics*. New York: Springer. ISBN 9813083263.
- Robert, C. P.; Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer. ISBN 0387212396.
- Rubinstein, R. Y.; Kroese, D. P. (2007). *Simulation and the Monte Carlo Method* (2nd ed.). New York: John Wiley & Sons. ISBN 9780470177938.
- Mosegaard, Klaus; Tarantola, Albert (1995). "Monte Carlo sampling of solutions to inverse problems". *J. Geophys. Res.* **100** (B7): 12431–12447. doi: 10.1029/94JB03097 (http://dx.doi.org/10.1029/94JB03097).
- Tarantola, Albert (2005). *http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/SIAM/index.html|Inverse Problem Theory*. Philadelphia: Society for Industrial and Applied Mathematics. ISBN 0898715725. http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/SIAM/index.html.

External links

- Overview and reference list (http://mathworld.wolfram.com/MonteCarloMethod.html), Mathworld
- Introduction to Monte Carlo Methods (http://www.ipp.mpg.de/de/for/bereiche/stellarator/Comp_sci/CompScience/csep/csep1.phy.ornl.gov/mc/mc.html), Computational Science Education Project
- Overview of formulas used in Monte Carlo simulation (http://www.sitmo.com/eqcat/15), the Quant Equation Archive, at sitmo.com
- The Basics of Monte Carlo Simulations (http://www.chem.unl.edu/zeng/joy/mclab/mcintro.html), University of Nebraska-Lincoln
- Introduction to Monte Carlo simulation (http://office.microsoft.com/en-us/assistance/HA011118931033.aspx) (for Excel), Wayne L. Winston
- Monte Carlo Methods - Overview and Concept (http://www.brighton-webs.co.uk/montecarlo/concept.asp), brighton-webs.co.uk
- Molecular Monte Carlo Intro (http://www.cooper.edu/engineering/chemechem/monte.html), Cooper Union
- Monte Carlo techniques applied in physics (http://homepages.ed.ac.uk/s0095122/Applet1-page.htm)
- MonteCarlo Simulation in Finance (http://www.global-derivatives.com/maths/k-o.php), global-derivatives.com

- Approximation of π with the Monte Carlo Method (http://twt.mpei.ac.ru/MAS/Worksheets/approxpi.mcd)
- Risk Analysis in Investment Appraisal (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=265905), The Application of Monte Carlo Methodology in Project Appraisal, Savvakis C. Savvides
- Probabilistic Assessment of Structures using the Monte Carlo method (http://en.wikiversity.org/wiki/Probabilistic_Assessment_of_Structures), Wikiuniversity paper for students of Structural Engineering
- A very intuitive and comprehensive introduction to Quasi-Monte Carlo methods (http://www.puc-rio.br/marco.ind/quasi_mc.html)
- Pricing using Monte Carlo simulation (http://knol.google.com/k/giancarlo-vercellino/pricing-using-monte-carlo-simulation/11d5i2rgd9gn5/3#), a practical example, Prof. Giancarlo Vercellino

Software

- The BUGS project (http://www.mrc-bsu.cam.ac.uk/bugs/) (including WinBUGS and OpenBUGS)
- Monte Carlo Simulation, Resampling, Bootstrap tool (http://www.statistics101.net)
- YASAI: Yet Another Simulation Add-In (http://yasai.rutgers.edu/) - Free Monte Carlo Simulation Add-In for Excel created by Rutgers University

Quantum Monte Carlo

Electronic structure methods
Tight binding
Nearly-free electron model
Hartree–Fock
Modern valence bond
Generalized valence bond
Møller–Plesset perturbation theory
Configuration interaction
Coupled cluster
Multi-configurational self-consistent field
Density functional theory
Quantum chemistry composite methods
Quantum Monte Carlo
k·p perturbation theory
Muffin-tin approximation
LCAO method

Quantum Monte Carlo is a large class of computer algorithms that simulate quantum systems with the idea of solving the many-body problem. They use, in one way or another, the Monte Carlo method to handle the many-dimensional integrals that arise. Quantum

Monte Carlo allows a direct representation of many-body effects in the wavefunction, at the cost of statistical uncertainty that can be reduced with more simulation time. For bosons, there exist numerically exact and polynomial-scaling algorithms. For fermions, there exist very good approximations and numerically exact exponentially scaling quantum Monte Carlo algorithms, but none that are both.

Background

In principle, any physical system can be described by the many-body Schrödinger equation as long as the constituent particles are not moving "too" fast; that is, they are not moving near the speed of light. This includes the electrons in almost every material in the world, so if we could solve the Schrödinger equation, we could predict the behavior of any electronic system, which has important applications in fields from computers to biology. This also includes the nuclei in Bose–Einstein condensate and superfluids such as liquid helium. The difficulty is that the Schrödinger equation involves a function of three times the number of particles and is difficult to solve even using parallel computing technology in a reasonable amount of time (less than 2 years). Traditionally, theorists have approximated the many-body wave function as an antisymmetric function of one-body orbitals, as shown concisely at this link.[1] This kind of formulation either limits the possible wave functions, as in the case of the Hartree-Fock (HF) approximation, or converges very slowly, as in configuration interaction. One of the reasons for the difficulty with an HF initial estimate (ground state seed, also known as Slater determinant) is that it is very difficult to model the electronic and nuclear cusps in the wavefunction. However, one does not generally model at this point of the approximation. As two particles approach each other, the wavefunction has exactly known derivatives.

Quantum Monte Carlo is a way around these problems because it allows us to model a many-body wavefunction of our choice directly. Specifically, we can use a Hartree-Fock approximation as our starting point but then multiplying it by any symmetric function, of which Jastrow functions are typical, designed to enforce the cusp conditions. Most methods aim at computing the ground-state wavefunction of the system, with the exception of path integral Monte Carlo and finite-temperature auxiliary field Monte Carlo, which calculate the density matrix.

There are several quantum Monte Carlo methods, each of which uses Monte Carlo in different ways to solve the many-body problem:

Quantum Monte Carlo methods

- Variational Monte Carlo : A good place to start; it is commonly used in many sorts of quantum problems.
- Diffusion Monte Carlo : The most common high-accuracy method for electrons (that is, chemical problems), since it comes quite close to the exact ground-state energy fairly efficiently. Also used for simulating the quantum behavior of atoms, etc.
- Path integral Monte Carlo : Finite-temperature technique mostly applied to bosons where temperature is very important, especially superfluid helium.
- Auxiliary field Monte Carlo : Usually applied to lattice problems, although there has been recent work on applying it to electrons in chemical systems.
- Reptation Monte Carlo : Recent zero-temperature method related to path integral Monte Carlo, with applications similar to diffusion Monte Carlo but with some different

tradeoffs.
- Gaussian quantum Monte Carlo

See also

- Stochastic Green Function (SGF) algorithm
- Monte Carlo method
- QMC@Home
- Quantum chemistry
- Density matrix renormalization group
- Time-evolving block decimation
- Metropolis algorithm
- Wavefunction optimization

References

[1] http://www.attaccalite.altervista.org/PhDThesis/html/node9.html

- Hammond, B.J.; W.A. Lester & P.J. Reynolds (1994) (in English). *http://www.worldscibooks.com/chemistry/1170.html|Monte Carlo Methods in Ab Initio Quantum Chemistry*. Singapore: World Scientific. ISBN 981-02-0321-7. OCLC 29594695 (http://worldcat.org/oclc/29594695). http://www.worldscibooks.com/chemistry/1170.html. Retrieved on 2007-01-18.
- Nightingale, M.P.; Umrigar, Cyrus J., ed (1999) (in English). *http://www.springer.com/west/0-7923-5552-0|Quantum Monte Carlo Methods in Physics and Chemistry*. Springer. ISBN 978-0-7923-5552-6. http://www.springer.com/west/0-7923-5552-0. Retrieved on 2007-01-18.
- W. M. C. Foulkes; L. Mitáš, R. J. Needs and G. Rajagopal (5 January 2001). "http://link.aps.org/abstract/RMP/v73/p33|Quantum Monte Carlo simulations of solids" (in English) (abstract). *Rev. Mod. Phys.* **73**: 33–83. doi: 10.1103/RevModPhys.73.33 (http://dx.doi.org/10.1103/RevModPhys.73.33). http://link.aps.org/abstract/RMP/v73/p33. Retrieved on 2007-01-18.
- Raimundo R. dos Santos (2003). "http://arxiv.org/abs/cond-mat/0303551v1|Introduction to Quantum Monte Carlo simulations for fermionic systems" (in English) (fulltext). *Braz. J. Phys.* **33**: 36. http://arxiv.org/abs/cond-mat/0303551v1. Retrieved on 2007-01-18.

External links

- QMCWIKI (http://www.qmcwiki.org/)
- Joint DEMOCRITOS-ICTP School on Continuum Quantum Monte Carlo Methods (http://cdsagenda5.ictp.trieste.it/full_display.php?ida=a0332&fid=)
- FreeScience Library -> Quantum Monte Carlo (http://freescience.info/books.php?id=35)
- UIUC 2007 Summer School on Computational Materials Science: Quantum Monte Carlo from Minerals and Materials to Molecules (http://www.mcc.uiuc.edu/summerschool/2007/qmc/)
- Quantum Monte Carlo in the Apuan Alps V (http://www.vallico.net/tti/tti.html) - international workshop, Vallico Sotto, Tuscany, 25 July-1 August 2009 (Click PUBLIC EVENTS) - Announcement (http://www.vallico.net/tti/qmcitaa_09/announcement.html), Poster (http://www.tcm.phy.cam.ac.uk/~mdt26/tti2/poster/

tti_c_poster_2009.png)
- Quantum Monte Carlo and the CASINO program IV (http://www.vallico.net/tti/tti.html) - summer school, Vallico Sotto, Tuscany, 2-9 August 2009 (Click PUBLIC EVENTS) - Announcement (http://www.vallico.net/tti/qmcatcp_09/announcement.html), Poster (http://www.tcm.phy.cam.ac.uk/~mdt26/tti2/poster/tti_ss_poster_2009.png)

Dynamics of Markovian particles

Dynamics of Markovian particles (or DMP) is the basis of a theory for kinetics of particles in open heterogeneous systems. It can be looked upon as an application of the notion of stochastic process conceived as a physical entity; e.g. the particle moves because there is a transition probability acting on it.

Two particular features of DMP might be noticed: (1) an ergodic like relation between the motion of particle and the corresponding steady state, and (2) the classic notion of geometric volume appears nowhere (e.g. a concept such as flow of "substance" is not expressed as liters per time unit but as number of particles per time unit). Though being primitive DMP has been applied for solving a classic paradox of the absorption of mercury by fish and by mollusks. The theory has also been applied for a purely probabilistic derivation of the fundamental physical principle: conservation of mass; this might be looked upon as a contribution to the old and ongoing discussion of the relation between physics and probability theory.

Sources

- Bergner---DMP, a kinetics of macroscopic particles in open heterogeneous systems [1]

References

[1] http://www.bergner.se/DMP/download.htm

Molecular Networks and Complex Molecule Dynamics

Metabolic network

A **metabolic network** is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. As such, these networks comprise the chemical reactions of metabolism as well as the regulatory interactions that guide these reactions.

With the sequencing of complete genomes, it is now possible to reconstruct the network of biochemical reactions in many organisms, from bacteria to human. Several of these networks are available online: Kyoto Encyclopedia of Genes and Genomes (KEGG)[1], EcoCyc [2] and BioCyc [3]. Metabolic networks are powerful tools, for studying and modelling metabolism. From the study of metabolic networks' topology with graph theory to predictive toxicology and ADME.

See also

- Metabolic network modelling
- Metabolic pathway

References

[1] http://www.genome.ad.jp
[2] http://www.ecocyc.org
[3] http://biocyc.org

Topological dynamics

In mathematics, **topological dynamics** is a branch of the theory of dynamical systems in which qualitative, asymptotic properties of dynamical systems are studied from the viewpoint of general topology.

Scope

The central object of study in topological dynamics is a **topological dynamical system**, i.e. a topological space, together with a continuous transformation, a continuous flow, or more generally, a semigroup of continuous transformations of that space. The origins of topological dynamics lie in the study of asymptotical properties of trajectories of systems of autonomous ordinary differential equations, in particular, the behavior of limit sets and various manifestations of "repetetiveness" of the motion, such as periodic trajectories, recurrence and minimality, stability, non-wandering points. George Birkhoff is considered to be the founder of the field. A structure theorem for minimal distal flows proved by Hillel Furstenberg in the early 1960s inspired much work on classification of minimal flows. A lot of research in the 1970s and 1980s was devoted to topological dynamics of one-dimensional maps, in particular, piecewise linear self-maps of the interval and the circle.

Unlike the theory of smooth dynamical systems, where the main object of study is a smooth manifold with a diffeomorphism or a smooth flow, phase spaces considered in topological dynamics are general metric spaces (usually, compact). This necessitates development of entirely different techniques but allows extra degree of flexibility even in the smooth setting, because invariant subsets of a manifold are frequently very complicated topologically (cf limit cycle, strange attractor); additionally, shift spaces arising via symbolic representations can be considered on an equal footing with more geometric actions. Topological dynamics has intimate connections with ergodic theory of dynamical systems, and many fundamental concepts of the latter have topological analogues (cf Kolmogorov–Sinai entropy and topological entropy).

See also

- Poincaré–Bendixson theorem
- Symbolic dynamics
- Topological conjugacy

References

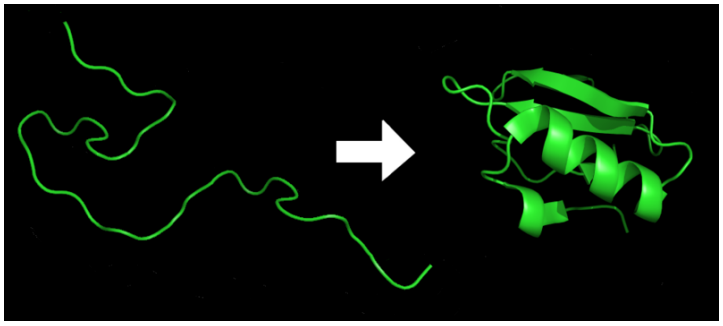
- D.V.Anosov (2001), http://eom.springer.de/T/t093030.htm|"Topological dynamics", in Hazewinkel, Michiel, *Encyclopaedia of Mathematics*, Kluwer Academic Publishers, ISBN 978-1556080104
- *Topological dynamics* [1] at Scholarpedia, curated by Joseph Auslander.
- Robert Ellis, *Lectures on topological dynamics*. W. A. Benjamin, Inc., New York 1969
- Walter Gottschalk, Gustav Hedlund, *Topological dynamics*. American Mathematical Society Colloquium Publications, Vol. 36. American Mathematical Society, Providence, R. I., 1955

- J. de Vries, *Elements of topological dynamics*. Mathematics and its Applications, 257. Kluwer Academic Publishers Group, Dordrecht, 1993 ISBN 0-7923-2287-8

References

[1] http://www.scholarpedia.org/article/Topological_dynamics

Protein folding



Protein before and after folding.

Protein folding is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil.[1] Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. This polypeptide lacks any developed three-dimensional structure (the left hand side of the neighboring figure). However amino acids interact with each other to produce a well-defined three dimensional structure, the folded protein (the right hand side of the figure), known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence.[2] .

For many proteins the correct three dimensional structure is essential to function.[3] Failure to fold into the intended shape usually produces inactive proteins with different properties including toxic prions. Several neurodegenerative and other diseases are believed to result from the accumulation of *misfolded* (incorrectly folded) proteins.[4]

Known facts about the process

The relationship between folding and amino acid sequence

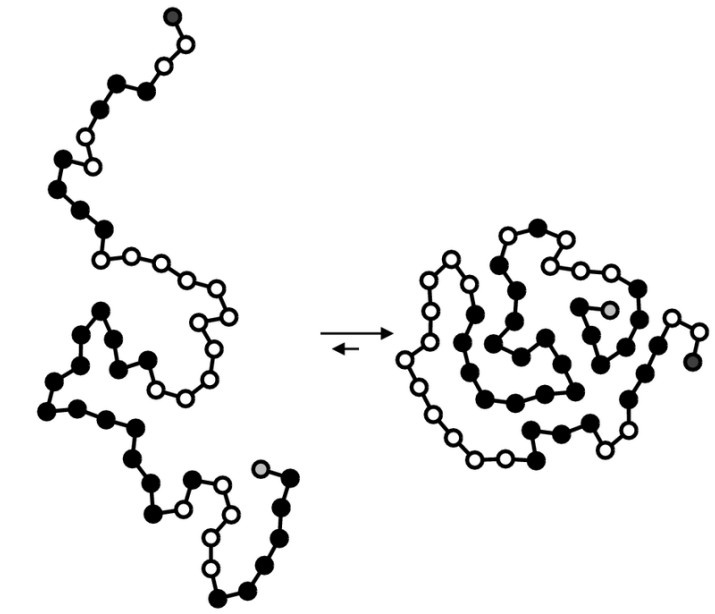


Illustration of the main driving force behind protein structure formation. In the compact fold (to the right), the hydrophobic amino acids (shown as black spheres) are in general shielded from the solvent.

The amino-acid sequence (or primary structure) of a protein defines its native conformation. A protein molecule folds spontaneously during or after synthesis. While these macromolecules may be regarded as "folding themselves", the process also depends on the solvent (water or lipid bilayer),[5] the concentration of salts, the temperature, and the presence of molecular chaperones.

Folded proteins usually have a hydrophobic core in which side chain packing stabilizes the folded state, and charged or polar side chains occupy the solvent-exposed surface where they interact with surrounding water. Minimizing the number of hydrophobic side-chains exposed to water is an important driving force behind the folding process,[6] . Formation of intramolecular hydrogen bonds provides another important contribution to protein stability.[7] The strength of hydrogen bonds depends on their environment, thus H-bonds enveloped in a hydrophobic core contribute more than H-bonds exposed to the aqueous environment to the stability of the native state.[8]

The process of folding *in vivo* often begins co-translationally, so that the N-terminus of the protein begins to fold while the C-terminal portion of the protein is still being synthesized by the ribosome. Specialized proteins called chaperones assist in the folding of other proteins.[9] A well studied example is the bacterial GroEL system, which assists in the folding of globular proteins. In eukaryotic organisms chaperones are known as heat shock proteins. Although most globular proteins are able to assume their native state unassisted, chaperone-assisted folding is often necessary in the crowded intracellular environment to prevent aggregation; chaperones are also used to prevent misfolding and aggregation which may occur as a consequence of exposure to heat or other changes in the cellular environment.

For the most part, scientists have been able to study many identical molecules folding together *en masse*. At the coarsest level, it appears that in transitioning to the native state, a given amino acid sequence takes on roughly the same route and proceeds through roughly the same intermediates and transition states. Often folding involves first the establishment of regular secondary and supersecondary structures, particularly alpha helices and beta sheets, and afterwards tertiary structure. Formation of quaternary

structure usually involves the "assembly" or "coassembly" of subunits that have already folded. The regular alpha helix and beta sheet structures fold rapidly because they are stabilized by intramolecular hydrogen bonds, as was first characterized by Linus Pauling. Protein folding may involve covalent bonding in the form of disulfide bridges formed between two cysteine residues or the formation of metal clusters. Shortly before settling into their more energetically favourable native conformation, molecules may pass through an intermediate "molten globule" state.

The essential fact of folding, however, remains that the amino acid sequence of each protein contains the information that specifies both the native structure and the pathway to attain that state. This is not to say that nearly identical amino acid sequences always fold similarly.[10] Conformations differ based on environmental factors as well; similar proteins fold differently based on where they are found. Folding is a spontaneous process independent of energy inputs from nucleoside triphosphates. The passage of the folded state is mainly guided by hydrophobic interactions, formation of intramolecular hydrogen bonds, and van der Waals forces, and it is opposed by conformational entropy.

Disruption of the native state

Under some conditions proteins will not fold into their biochemically functional forms. Temperatures above or below the range that cells tend to live in will cause thermally unstable proteins to unfold or "denature" (this is why boiling makes an egg white turn opaque). High concentrations of solutes, extremes of pH, mechanical forces, and the presence of chemical denaturants can do the same. A fully denatured protein lacks both tertiary and secondary structure, and exists as a so-called random coil. Under certain conditions some proteins can refold; however, in many cases denaturation is irreversible.[11] Cells sometimes protect their proteins against the denaturing influence of heat with enzymes known as chaperones or heat shock proteins, which assist other proteins both in folding and in remaining folded. Some proteins never fold in cells at all except with the assistance of chaperone molecules, which either isolate individual proteins so that their folding is not interrupted by interactions with other proteins or help to unfold misfolded proteins, giving them a second chance to refold properly. This function is crucial to prevent the risk of precipitation into insoluble amorphous aggregates.

Incorrect protein folding and neurodegenerative disease

Aggregated proteins are associated with prion-related illnesses such as Creutzfeldt-Jakob disease, bovine spongiform encephalopathy (mad cow disease), amyloid-related illnesses such as Alzheimer's Disease and familial amyloid cardiomyopathy or polyneuropathy, as well as intracytoplasmic aggregation diseases such as Huntington's and Parkinson's disease. These age onset degenerative diseases are associated with the multimerization of misfolded proteins into insoluble, extracellular aggregates and/or intracellular inclusions including cross-beta sheet amyloid fibrils; it is not clear whether the aggregates are the cause or merely a reflection of the loss of protein homeostasis, the balance between synthesis, folding, aggregation and protein turnover. Misfolding and excessive degradation instead of folding and function leads to a number of proteopathy diseases such as antitrypsin-associated Emphysema, cystic fibrosis and the lysosomal storage diseases, where loss of function is the origin of the disorder. While protein replacement therapy has historically been used to correct the latter disorders, an emerging approach is to use pharmaceutical chaperones to fold mutated proteins to render them functional. Chris

Dobson, Jeffery W. Kelly, Dennis Selkoe, Stanley Prusiner, Peter T. Lansbury, William E. Balch, Richard I. Morimoto, Susan L. Lindquist and Byron C. Caughey have all contributed to this emerging understanding of protein-misfolding diseases.

Kinetics and the Levinthal Paradox

The duration of the folding process varies dramatically depending on the protein of interest. When studied outside the cell, the slowest folding proteins require many minutes or hours to fold primarily due to proline isomerization, and must pass through a number of intermediate states, like checkpoints, before the process is complete.[12] On the other hand, very small single-domain proteins with lengths of up to a hundred amino acids typically fold in a single step.[13] Time scales of milliseconds are the norm and the very fastest known protein folding reactions are complete within a few microseconds.[14]

The Levinthal paradox[15] observes that if a protein were to fold by sequentially sampling all possible conformations, it would take an astronomical amount of time to do so, even if the conformations were sampled at a rapid rate (on the nanosecond or picosecond scale). Based upon the observation that proteins fold much faster than this, Levinthal then proposed that a random conformational search does not occur, and the protein must, therefore, fold through a series of meta-stable intermediate states.

Techniques for studying protein folding

Circular Dichroism

Circular dichroism is one of the most general and basic tools to study protein folding. Circular dichroism spectroscopy measures the absorption of circularly polarized light. In proteins, structures such as alpha helicies and beta sheets are chiral, and thus absorb such light. The absorption of this light acts as a marker of the degree of foldedness of the protein ensemble. This technique can be used to measure equilibrium unfolding of the protein by measuring the change in this absorption as a function of denaturant concentration or temperature. A denaturant melt measures the free energy of unfolding as well as the protein's m value, or denaturant dependence. A temperature melt measures the melting temperature (T_m) of the protein. This type of spectroscopy can also be combined with fast-mixing devices, such as stopped flow, to measure protein folding kinetics and to generate chevron plots.

Dual Polarisation Interferometry

Dual Polarisation Interferometry is a relatively new benchtop technique for measuring the overall change in protein size and fold density during interactions or other stimulus. The technique captures a layer of protein on a glass slide and, using two polarisations of light, measures the conformation and conformational changes with a time resolution of circa 10Hz at a dimensional resolution of 0.01nm. The method is quantitative and can be compared directly to what one would expect of crystallography data.

Vibrational circular dichroism of proteins

The more recent developments of vibrational circular dichroism (VCD) techniques for proteins, currently involving Fourier transform (FFT) instruments, provide powerful means for determining protein conformations in solution even for very large protein molecules. Such VCD studies of proteins are often combined with X-ray diffraction of protein crystals, FT-IR data for protein solutions in heavy water (D_2O), or *ab initio* quantum computations to provide unambiguous structural assignments that are unobtainable from CD.

Modern studies of folding with high time resolution

The study of protein folding has been greatly advanced in recent years by the development of fast, time-resolved techniques. These are experimental methods for rapidly triggering the folding of a sample of unfolded protein, and then observing the resulting dynamics. Fast techniques in widespread use include neutron scattering[16] , ultrafast mixing of solutions, photochemical methods, and laser temperature jump spectroscopy. Among the many scientists who have contributed to the development of these techniques are Jeremy Cook, Heinrich Roder, Harry Gray, Martin Gruebele, Brian Dyer, William Eaton, Sheena Radford, Chris Dobson, Sir Alan R. Fersht and Bengt Nölting.

Energy landscape theory of protein folding

The protein folding phenomenon was largely an experimental endeavor until the formulation of energy landscape theory by Joseph Bryngelson and Peter Wolynes in the late 1980s and early 1990s. This approach introduced the principle of minimal frustration, which asserts that evolution has selected the amino acid sequences of natural proteins so that interactions between side chains largely favor the molecule's acquisition of the folded state. Interactions that do not favor folding are selected against, although some residual *frustration* is expected to exist. A consequence of these evolutionarily selected sequences is that proteins are generally thought to have globally "funneled energy landscapes" (coined by José Onuchic[reference needed]) that are largely directed towards the native state. This "folding funnel" landscape allows the protein to fold to the native state through any of a large number of pathways and intermediates, rather than being restricted to a single mechanism. The theory is supported by both computational simulations of model proteins and numerous experimental studies, and it has been used to improve methods for protein structure prediction and design.

Computational prediction of protein tertiary structure

De novo or *ab initio* techniques for computational protein structure prediction is related to, but strictly distinct from, studies involving protein folding. Molecular Dynamics (MD) is an important tool for studying protein folding and dynamics in silico. Because of computational cost, ab initio MD folding simulations with explicit water are limited to peptides and very small proteins. MD simulations of larger proteins remain restricted to dynamics of the experimental structure or its high-temperature unfolding. In order to simulate long time folding processes (beyond about 1 microsecond), like folding of small-size proteins (about 50 residues) or larger, some approximations or simplifications in protein models need to be introduced. An approach using reduced protein representation (pseudo-atoms representing groups of atoms are defined) and statistical potential is not only useful in protein structure prediction, but is also capable of reproducing the folding pathways.[17]

There are distributed computing projects which use idle CPU time of personal computers to solve problems such as protein folding or prediction of protein structure. People can run these programs on their computer or PlayStation 3 to support them. See links below (for example Folding@Home) to get information about how to participate in these projects.

Experimental techniques of protein structure determination

Folded structures of proteins are routinely determined by X-ray crystallography and NMR.

See also

- Anfinsen's dogma
- Chevron plot
- Denaturation (biochemistry)
- Denaturation midpoint
- Downhill folding
- Equilibrium unfolding
- Folding (chemistry)
- Folding@Home
- Foldit computer game
- Levinthal paradox
- Protein design
- Protein dynamics
- Protein structure prediction
- Protein structure prediction software
- Rosetta@Home
- Software for molecular mechanics modeling

References

[1] Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters (2002). http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=mboc4[book]+AND+37 Shape and Structure of Proteins". *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science. ISBN 0-8153-3218-1.

[2] Anfinsen C (1972). "The formation and stabilization of protein structure". *Biochem. J.* **128** (4): 737–49. PMID 4565129.

[3] Jeremy M. Berg, John L. Tymoczko, Lubert Stryer; Web content by Neil D. Clarke (2002). http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=stryer[book]+AND+215 Protein Structure and Function". *Biochemistry*. San Francisco: W.H. Freeman. ISBN 0-7167-4684-0.

[4] http://folding.stanford.edu/science.html|"Science of Folding@Home". July 18, 2005. http://folding.stanford.edu/science.html. Retrieved on 2007-04-22.

[5] van den Berg B, Wain R, Dobson CM, Ellis RJ (August 2000). "http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=306593|Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell". *Embo J.* **19** (15): 3870–5. doi: 10.1093/emboj/19.15.3870 (http://dx.doi.org/10.1093/emboj/19.15.3870). PMID 10921869.

[6] Pace C, Shirley B, McNutt M, Gajiwala K (01 Jan 1996). "http://www.fasebj.org/cgi/reprint/10/1/75|Forces contributing to the conformational stability of proteins". *Faseb J.* **10** (1): 75–83. PMID 8566551. http://www.fasebj.org/cgi/reprint/10/1/75.

[7] Rose G, Fleming P, Banavar J, Maritan A (2006). "A backbone-based theory of protein folding". *Proc. Natl. Acad. Sci. U.S.A.* **103** (45): 16623–33. doi: 10.1073/pnas.0606843103 (http://dx.doi.org/10.1073/pnas.0606843103). PMID 17075053.

[8] Deechongkit S, Nguyen H, Dawson PE, Gruebele M, Kelly JW (2004). "Context Dependent Contributions of Backbone H-Bonding to β-Sheet Folding Energetics". *Nature* **403** (45): 101–5. doi: 10.1073/pnas.0606843103 (http://dx.doi.org/10.1073/pnas.0606843103). PMID 17075053.

[9] Lee S, Tsai F (2005). "http://www.jbmb.or.kr/fulltext/jbmb/view.php?vol=38&page=259|Molecular chaperones in protein quality control". *J. Biochem. Mol. Biol.* **38** (3): 259–65. PMID 15943899. http://www.jbmb.or.kr/fulltext/jbmb/view.php?vol=38&page=259.

[10] Alexander PA, He Y, Chen Y, Orban J, Bryan PN. (2007). "http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=1906725|The design and characterization of two proteins with 88% sequence identity but different structure and function". *Proc Natl Acad Sci U S A.* **104** (29): 11963–8. doi: 10.1073/pnas.0700922104 (http://dx.doi.org/10.1073/pnas.0700922104). PMID 17609385.

[11] Shortle D (01 Jan 1996). "http://www.fasebj.org/cgi/reprint/10/1/27|The denatured state (the other half of the folding equation) and its role in protein stability". *Faseb J.* **10** (1): 27–34. PMID 8566543. http://www.fasebj.org/cgi/reprint/10/1/27.

[12] Kim PS, Baldwin RL (1990). "Intermediates in the folding reactions of small proteins". *Annu. Rev. Biochem.* **59**: 631–60. doi: 10.1146/annurev.bi.59.070190.003215 (http://dx.doi.org/10.1146/annurev.bi.59.070190.003215). PMID 2197986.

[13] Jackson SE (August 1998). "http://biomednet.com/elecref/13590278003R0081|How do small single-domain proteins fold?". *Fold Des* **3** (4): R81–91. doi: 10.1016/S1359-0278(98)00033-9 (http://dx.doi.org/10.1016/S1359-0278(98)00033-9). PMID 9710577. http://biomednet.com/elecref/13590278003R0081.

[14] Kubelka J, Hofrichter J, Eaton WA (February 2004). "The protein folding 'speed limit'". *Curr. Opin. Struct. Biol.* **14** (1): 76–88. doi: 10.1016/j.sbi.2004.01.013 (http://dx.doi.org/10.1016/j.sbi.2004.01.013). PMID 15102453.

[15] C. Levinthal (1968). "http://www.biochem.wisc.edu/courses/biochem704/Reading/Levinthal1968.pdf|Are there pathways for protein folding?". *J. Chim. Phys.* **65**: 44–5. http://www.biochem.wisc.edu/courses/biochem704/Reading/Levinthal1968.pdf.

[16] Bu Z, Cook J,Callaway DJE (2001). "Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbuminC". *J Mol Biol* **312** (4): 865-873. doi: 10.1006/jmbi.2001.5006 (http://dx.doi.org/10.1006/jmbi.2001.5006).

[17] Kmiecik S and Kolinski A (2007). "Characterization of protein-folding pathways by reduced-space modeling". *Proc. Natl. Acad. Sci. U.S.A.* **104** (30): 12330–5. doi: 10.1073/pnas.0702265104 (http://dx.doi.org/10.1073/pnas.0702265104). PMID 17636132.

External links

- FoldIt - Folding Protein Game (http://fold.it/portal/info/science)
- Folding@Home (http://www.stanford.edu/group/pandegroup/folding/about.html)
- Rosetta@Home (http://boinc.bakerlab.org/rosetta)

Protein-protein interaction

Protein-protein interactions involve not only the direct-contact association of protein molecules but also longer range interactions through the electrolyte, aqueous solution medium surrounding neighbor hydrated proteins over distances from less than one nanometer to distances of several tens of nanometers. Furthermore, such protein-protein interactions are thermodynamically linked functions[1] of dynamically bound ions and water that exchange rapidly with the surrounding solution by comparison with the molecular tumbling rate (or correlation times) of the interacting proteins. Protein associations are also studied from the perspectives of biochemistry, quantum chemistry, molecular dynamics, signal transduction and other metabolic or genetic/epigenetic networks. Indeed, protein-protein interactions are at the core of the entire Interactomics system of any living cell.

The interactions between proteins are important for very numerous—if not all—biological functions. For example, signals from the exterior of a cell are mediated to the inside of that cell by protein-protein interactions of the signaling molecules. This process, called signal transduction, plays a fundamental role in many biological processes and in many diseases (e.g. cancers). Proteins might interact for a long time to form part of a protein complex, a protein may be carrying another protein (for example, from cytoplasm to nucleus or vice versa in the case of the nuclear pore importins), or a protein may interact briefly with another protein just to modify it (for example, a protein kinase will add a phosphate to a target protein). This modification of proteins can itself change protein-protein interactions. For example, some proteins with SH2 domains only bind to other proteins when they are phosphorylated on the amino acid tyrosine while bromodomains specifically recognise acetylated lysines. In conclusion, protein-protein interactions are of central importance for virtually every process in a living cell. Information about these interactions improves our understanding of diseases and can provide the basis for new therapeutic approaches.

Methods to investigate protein-protein interactions

Biochemical methods

As protein-protein interactions are so important there are a multitude of methods to detect them. Each of the approaches has its own strengths and weaknesses, especially with regard to the sensitivity and specificity of the method. A high sensitivity means that many of the interactions that occur in reality are detected by the screen. A high specificity indicates that most of the interactions detected by the screen are also occurring in reality.

- Co-immunoprecipitation is considered to be the gold standard assay for protein-protein interactions, especially when it is performed with endogenous (not overexpressed and not tagged) proteins. The protein of interest is isolated with a specific antibody. Interaction partners which stick to this protein are subsequently identified by western blotting. Interactions detected by this approach are considered to be real. However, this method can only verify interactions between suspected interaction partners. Thus, it is not a screening approach. A note of caution also is that immunoprecipitation experiments reveal direct and indirect interactions. Thus, positive results may indicate that two proteins interact directly or may interact via a bridging protein.

- Bimolecular Fluorescence Complementation (BiFC) is a new technique in observing the interactions of proteins. Combining with other new techniques, this method can be used to screen protein-protein interactions and their modulators [2] .
- Affinity electrophoresis as used for estimation of binding constants, as for instance in lectin affinity electrophoresis or characterization of molecules with specific features like glycan content or ligand binding.
- Pull-down assays are a common variation of immunoprecipitation and immunoelectrophoresis and are used identically, although this approach is more amenable to an initial screen for interacting proteins.
- Label transfer can be used for screening or confirmation of protein interactions and can provide information about the interface where the interaction takes place. Label transfer can also detect weak or transient interactions that are difficult to capture using other *in vitro* detection strategies. In a label transfer reaction, a known protein is tagged with a detectable label. The label is then passed to an interacting protein, which can then be identified by the presence of the label.
- The yeast two-hybrid screen investigates the interaction between artificial fusion proteins inside the nucleus of yeast. This approach can identify binding partners of a protein in an unbiased manner. However, the method has a notorious high false-positive rate which makes it necessary to verify the identified interactions by co-immunoprecipitation.
- *In-vivo* crosslinking of protein complexes using photo-reactive amino acid analogs was introduced in 2005 by researchers from the Max Planck Institute [3] In this method, cells are grown with photoreactive diazirine analogs to leucine and methionine, which are incorporated into proteins. Upon exposure to ultraviolet light, the diazirines are activated and bind to interacting proteins that are within a few angstroms of the photo-reactive amino acid analog.
- Tandem affinity purification (TAP) method allows high throughput identification of protein interactions. In contrast to Y2H approach accuracy of the method can be compared to those of small-scale experiments (Collins et al., 2007) and the interactions are detected within the correct cellular environment as by co-immunoprecipitation. However, the TAP tag method requires two successive steps of protein purification and consequently it can not readily detect transient protein-protein interactions. Recent genome-wide TAP experiments were performed by Krogan et al., 2006 and Gavin et al., 2006 providing updated protein interaction data for yeast organism.
- Chemical crosslinking is often used to "fix" protein interactions in place before trying to isolate/identify interacting proteins. Common crosslinkers for this application include the non-cleavable NHS-ester crosslinker, *bis*-sulfosuccinimidyl suberate (BS3); a cleavable version of BS3, dithiobis(sulfosuccinimidyl propionate) (DTSSP); and the imidoester crosslinker dimethyl dithiobispropionimidate (DTBP) that is popular for fixing interactions in ChIP assays.
- Chemical crosslinking followed by high mass MALDI mass spectrometry can be used to analyze intact protein interactions in place before trying to isolate/identify interacting proteins. This method detects interactions among non-tagged proteins and is available from CovalX.
- SPINE (Strep-protein interaction experiment) [4] uses a combination of reversible crosslinking with formaldehyde and an incorporation of an affinity tag to detect interaction partners *in vivo*.

- Quantitative immunoprecipitation combined with knock-down (QUICK) relies on co-immunoprecipitation, quantitative mass spectrometry (SILAC) and RNA interference (RNAi). This method detects interactions among endogenous non-tagged proteins[5] . Thus, it has the same high confidence as co-immunoprecipitation. However, this method also depends on the availability of suitable antibodies.

Physical/Biophysical and Theoretical methods

- Dual Polarisation Interferometry (DPI) can be used to measure protein-protein interactions. DPI provides real-time, high-resolution measurements of molecular size, density and mass. While tagging is not necessary, one of the protein species must be immobilized on the surface of a waveguide.
- Static Light scattering (SLS) measures changes in the Rayleigh scattering of protein complexes in solution and can non-destructively characterize both weak and strong interactions without tagging or immobilization of the protein. The measurement consists of mixing a series of aliquots of different concentrations or compositions with the anylate, measuring the effect of the changes in light scattering as a result of the interaction, and fitting the correlated light scattering changes with concentration to a model. Weak, non-specific interactions are typically characterized via the second virial coefficient. This type of analysis can determine the equilibrium association constant for associated complexes.[6] . Additional light scattering methods for protein activity determination were previously developed by Timasheff. More recent Dynamic Light scattering (DLS) methods for proteins were reported by H. Chou that are also applicable at high protein concentrations and in protein gels; DLS may thus also be applicable for *in vivo* cytoplasmic observations of various protein-protein interactions.
- Surface plasmon resonance can be used to measure protein-protein interaction.
- With Fluorescence correlation spectroscopy, one protein is labeled with a fluorescent dye and the other is left unlabeled. The two proteins are then mixed and the data outputs the fraction of the labeled protein that is unbound and bound to the other protein, allowing you to get a measure of K_D and binding affinity. You can also take time-course measurements to characterize binding kinetics. FCS also tells you the size of the formed complexes so you can measure the stoichiometry of binding. A more powerful methods is [[fluorescence cross-correlation spectroscopy (FCCS) that employs double labeling techniques and cross-correlation resulting in vastly improved signal-to-noise ratios over FCS. Furthermore, the two-photon and three-photon excitation practically eliminates photobleaching effects and provide ultra-fast recording of FCCS or FCS data.
- Fluorescence resonance energy transfer (FRET) is a common technique when observing the interactions of only two different proteins[7] .
- Protein activity determination by NMR multi-nuclear relaxation measurements, or 2D-FT NMR spectroscopy in solutions, combined with nonlinear regression analysis of NMR relaxation or 2D-FT spectroscopy data sets. Whereas the concept of water activity is widely known and utilized in the applied biosciences, its complement--the protein activity which quantitates protein-protein interactions-- is much less familiar to bioscientists as it is more difficult to determine in dilute solutions of proteins; protein activity is also much harder to determine for concentrated protein solutions when protein aggregation, not merely transient protein association, is often the dominant process[8] .
- Theoretical modeling of protein-protein interactions involves a detailed physical chemistry/thermodynamic understanding of several effects involved, such as

intermolecular forces, ion-binding, proton fluctuations and proton exchange. The theory of thermodynamically linked functions is one such example in which ion-binding and protein-protein interactions are treated as linked processes; this treatment is especially important for proteins that have enzymatic activity which depends on cofactor ions dynamically bound at the enzyme active site, as for example, in the case of oxygen-evolving enzyme system (OES) in photosythetic biosystems where the oxygen molecule binding is linked to the chloride anion binding as well as the linked state transition of the manganese ions present at the active site in Photosystem II(PSII). Another example of thermodynamically linked functions of ions and protein activity is that of divalent calcium and magnesium cations to myosin in mechanical energy transduction in muscle. Last-but-not least, chloride ion and oxygen binding to hemoglobin (from several mammalian sources, including human) is a very well-known example of such thermodynamically linked functions for which a detailed and precise theory has been already developed.

- Molecular dynamics (MD) computations of protein-protein interactions.
- Protein-protein docking, the prediction of protein-protein interactions based only on the three-dimensional protein structures from X-ray diffraction of protein crystals might not be satisfactory.[9] [10]

Network visualization of protein-protein interactions

Visualization of protein-protein interaction networks is a popular application of scientific visualization techniques. Although protein interaction diagrams are common in textbooks, diagrams of whole cell protein interaction networks were not as common since the level of complexity made them difficult to generate. One example of a manually produced molecular interaction map is Kurt Kohn's 1999 map of cell cycle control.[11] Drawing on Kohn's map, in 2000 Schwikowski, Uetz, and Fields published a paper on protein-protein interactions in yeast, linking together 1,548 interacting proteins determined by two-hybrid testing. They used a force-directed (Sugiyama) graph drawing algorithm to automatically generate an image of their network.[12] [13] [14] .

An experimental view of Kurt Kohn's 1999 map gmap [15]. Image was merged via gimp 2.2.17 and then uploaded to maplib.net

See also

- Interactomics
- Signal transduction
- Biophysical techniques
- Biochemistry methods
- Genomics
- Complex systems biology
- Complex systems
- Immunoprecipitation
- Protein-protein interaction prediction
- Protein-protein interaction screening
- BioGRID, a public repository for protein and genetic interactions
- Database of Interacting Proteins (DIP)
- NCIBI National Center for Integrative Biomedical Informatics

- Biotechnology
- Protein nuclear magnetic resonance spectroscopy
- 2D-FT NMRI and Spectroscopy
- Fluorescence correlation spectroscopy
- Fluorescence cross-correlation spectroscopy
- Light scattering
- ConsensusPathDB

References

[1] Kinetic Linked-Function Analysis of the Multiligand Interactions on Mg2+-Activated Yeast Pyruvate Kinase. Thomas J. Bollenbach and Thomas Nowak., *Biochemistry*, 2001, 40 (43), pp. 13097–13106

[2] Lu JP, Beatty LK, Pinthus JH. (2008). "Dual expression recombinase based (DERB) single vector system for high throughput screening and verification of protein interactions in living cells.". *Nature Precedings <http://hdl.handle.net/10101/npre.2008.1550.2>*.

[3] Suchanek, M., Radzikowska, A., and Thiele, C. (2005). "Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells". *Nature Methods* **2**: 261–268. doi: 10.1038/nmeth752 (http://dx.doi.org/10.1038/nmeth752). PMID 15782218.

[4] Herzberg C., Weidinger LA., Dörrbecker B., Hübner S., Stülke J. and Commichau FM. (2007). "SPINE: A method for the rapid detection and analysis of protein-protein interactions *in vivo*". *Proteomics* **7(22)**: 4032–4035. doi: 10.1002/pmic.200700491 (http://dx.doi.org/10.1002/pmic.200700491). PMID 17994626.

[5] Selbach, M., Mann, M. (2006). "Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK)". *Nature Methods* **3**: 981–983. doi: 10.1038/nmeth972 (http://dx.doi.org/10.1038/nmeth972). PMID 17072306.

[6] Arun K. Attri and Allen P. Minton (2005). "Composition gradient static light scattering: A new technique for rapid detection and quantitative characterization of reversible macromolecular hetero-associations in solution". *Analytical Biochemistry* **346**: 132–138. doi: 10.1016/j.ab.2005.08.013 (http://dx.doi.org/10.1016/j.ab.2005.08.013). PMID 16188220.

[7] Gadella TW Jr., *FRET and FLIM techniques*, 33. Imprint: Elsevier, ISBN 978-0-08-054958-3. (2008) 560 pages.

[8] #Baianu, I.C.; Kumosinski, Thomas (August 1993). "NMR Principles and Applications to Protein Structure, Activity and Hydration.,". *Ch.9 in* Physical Chemistry of Food Processes: Advanced Techniques and Applications. (New York: Van Nostrand-Reinhold) **2**: 338–420. ISBN 0-442-00582-2.

[9] Bonvin AM (2006). "Flexible protein-protein docking". *Current Opinion in Structural Biology* **16**: 194–200. doi: 10.1016/j.sbi.2006.02.002 (http://dx.doi.org/10.1016/j.sbi.2006.02.002). PMID 16488145.

[10] Gray JJ (2006). "High-resolution protein-protein docking". *Current Opinion in Structural Biology* **16**: 183–193. doi: 10.1016/j.sbi.2006.03.003 (http://dx.doi.org/10.1016/j.sbi.2006.03.003). PMID 16546374.

[11] Kurt W. Kohn (1999). "http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=10436023|Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems". *Molecular Biology of the Cell* **10** (8): 2703–2734. PMID 10436023. http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=10436023.

[12] Benno Schwikowski1, Peter Uetz, and Stanley Fields (2000). "http://igtmv1.fzk.de/www/itg/uetz/publications/Schwikowski2000.pdf|A network of protein−protein interactions in yeast". *Nature Biotechnology* **18**: 1257–1261. doi: 10.1038/82360 (http://dx.doi.org/10.1038/82360). PMID 11101803. http://igtmv1.fzk.de/www/itg/uetz/publications/Schwikowski2000.pdf.

[13] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 17:1030-2.

[14] Prieto C, De Las Rivas J (2006). APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res. 34:W298-302.

[15] http://www.maplib.net/map.php?id=1700&lat=-52.67138590320257&lng=34.3817138671875&z=9

Further reading

1. Gadella TW Jr., *FRET and FLIM techniques*, 33. Imprint: Elsevier, ISBN 978-0-08-054958-3. (2008) 560 pages
2. Langel FD, et al., Multiple protein domains mediate interaction between Bcl10 and Malt1, *J. Biol. Chem.*, (2008) 283(47):32419-31
3. Clayton AH. , The polarized AB plot for the frequency-domain analysis and representation of fluorophore rotation and resonance energy homotransfer. J *Microscopy*. (2008) 232(2):306-12
4. Clayton AH, et al., Predominance of activated EGFR higher-order oligomers on the cell surface. *Growth Factors* (2008) 20:1
5. Plowman et al., Electrostatic Interactions Positively Regulate K-Ras Nanocluster Formation and Function. *Molecular and Cellular Biology* (2008) 4377–4385
6. Belanis L, et al., Galectin-1 Is a Novel Structural Component and a Major Regulator of H-Ras Nanoclusters. *Molecular Biology of the Cell* (2008) 19:1404–1414
7. Van Manen HJ, Refractive index sensing of green fluorescent proteins in living cells using fluorescence lifetime imaging microscopy. Biophys J. (2008) 94(8):L67-9
8. Van der Krogt GNM, et al., A Comparison of Donor-Acceptor Pairs for Genetically Encoded FRET Sensors: Application to the Epac cAMP Sensor as an Example, PLoS ONE, (2008) 3(4):e1916
9. Dai X, et al., Fluorescence intensity and lifetime imaging of free and micellar-encapsulated doxorubicin in living cells. *Nanomedicine*. (2008) **4**(1):49-56.
10. Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, BioScience (Ed. Klinge & Owman) p.180.
11. Near Infrared Microspectroscopy, Fluorescence Microspectroscopy, Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In Oil Extraction and Analysis., D. Luthria, Editor pp.241–273, AOCS Press., Champaign, IL
12. Richard R. Ernst. 1992. Nuclear Magnetic Resonance Fourier Transform (2D-FT) Spectroscopy.Nobel Lecture, on December 9, 1992.
13. Baianu, I.C.; Kumosinski, Thomas (August 1993). "NMR Principles and Applications to Protein Structure, Activity and Hydration.,". *Ch.9 in* Physical Chemistry of Food Processes: Advanced Techniques and Applications. (New York: Van Nostrand-Reinhold) **2**: 338–420. ISBN 0-442-00582-2.
14. Kurt Wüthrich in 1982-1986 : 2D-FT NMR of solutions (http://en.wikipedia.org/wiki/Nuclear_magnetic_resonance#Nuclear_spin_and_magnets)
15. Charles P. Slichter.1996. *Principles of Magnetic Resonance.*, Springer: Berlin and New York, Third Edition., 651pp. ISBN 0-387-50157-6.
16. Kurt Wüthrich. *Protein structure determination in solution by NMR spectroscopy* . J Biol Chem. 1990, December 25;265(36):22059-62.

External links

- National Center for Integrative Biomedical Informatics (NCIBI) (http://portal.ncibi.org/gateway/)
- Proteins and Enzymes (http://www.dmoz.org/Science/Biology/Biochemistry_and_Molecular_Biology/Biomolecules/Proteins_and_Enzymes/) at the Open Directory Project
- FLIM Applications (http://www.nikoninstruments.com/infocenter.php?n=FLIM) FLIM is also often used in microspectroscopic/ chemical imaging, or microscopic, studies to monitor spatial and temporal protein-protein interactions, properties of membranes and interactions with nucleic acids in living cells.
- Arabidopsis thaliana protein interaction network (http://bioinfo.esalq.usp.br/atpin)

DNA Dynamics

DNA Molecular dynamics modeling involves simulations of DNA molecular geometry and topology changes with time as a result of both intra- and inter- molecular interactions of DNA. Whereas molecular models of Deoxyribonucleic acid (DNA) molecules such as closely packed spheres (CPK models) made of plastic or metal wires for 'skeletal models' are useful representations of static DNA structures, their usefulness is very limited for representing complex DNA dynamics. Computer molecular modeling allows both animations and molecular dynamics simulations that are very important for understanding how DNA functions *in vivo*.

An old standing dynamic problem is how DNA "self-replication" takes place in living cells that should involve transient uncoiling of supercoiled DNA fibers. Although DNA consists of relatively rigid, very large elongated biopolymer molecules called "fibers" or chains its molecular structure *in vivo* undergoes dynamic configuration changes that involve dynamically attached water molecules, ions or proteins/enzymes. Supercoiling, packing with histones in chromosome structures, and other such supramolecular aspects also involve *in vivo* DNA topology which is even more complex than DNA molecular geometry, thus turning molecular modeling of DNA dynamics into a series of challenging problems for biophysical chemists, molecular biologists and biotechnologists. Thus, DNA exists in multiple stable geometries (called conformational isomerism) and has a rather large number of configurational, quantum states which are close to each other in energy on the potential energy surface of the DNA molecule.

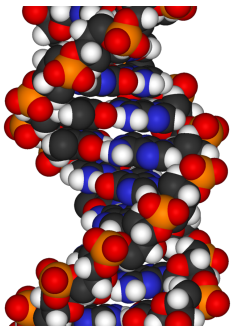
Such varying molecular geometries can also be computed, at least in principle, by employing *ab initio* quantum chemistry methods that can attain high accuracy for small molecules, although claims that acceptable accuracy can be also achieved for polynuclelotides, as well as DNA conformations, were recently made on the basis of VCD spectral data. Such quantum geometries define an important class of *ab initio* molecular models of DNA whose exploration has barely started especially in connection with results obtained by VCD in solutions. More detailed comparisons with such *ab initio* quantum computations are in principle obtainable through 2D-FT NMR spectroscopy and relaxation studies of polynucleotide solutions or specifically labeled DNA, as for example with deuterium labels.

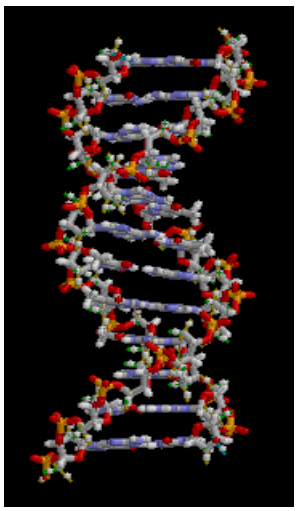
Importance of DNA molecular structure and dynamics modeling for Genomics and beyond

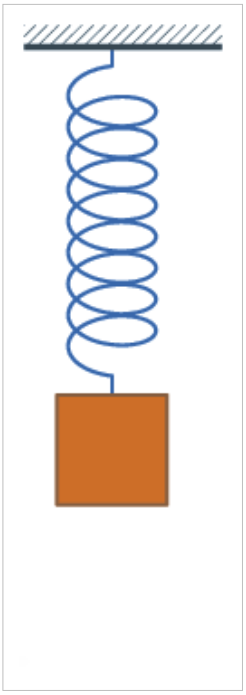
From the very early stages of structural studies of DNA by X-ray diffraction and biochemical means, molecular models such as the Watson-Crick double-helix model were successfully employed to solve the 'puzzle' of DNA structure, and also find how the latter relates to its key functions in living cells. The first high quality X-ray diffraction patterns of A-DNA were reported by Rosalind Franklin and Raymond Gosling in 1953[1] . The first reports of a double-helix molecular model of B-DNA structure were made by Watson and Crick in 1953[2] [3] . Then Maurice F. Wilkins, A. Stokes and H.R. Wilson, reported the first X-ray patterns of *in vivo* B-DNA in partially oriented salmon sperm heads [4] . The development of the first correct double-helix molecular model of DNA by Crick and Watson may not have been possible without the biochemical evidence for the nucleotide base-pairing ([A---T]; [C---G]), or Chargaff's rules[5] [6] [7] [8] [9] [10] . Although such initial studies of DNA structures with the help of molecular models were essentially static, their consequences for explaining the *in vivo* functions of DNA were significant in the areas of protein biosynthesis and the quasi-universality of the genetic code. Epigenetic transformation studies of DNA *in vivo* were however much slower to develop in spite of their importance for embryology, morphogenesis and cancer research. Such chemical dynamics and biochemical reactions of DNA are much more complex than the molecular dynamics of DNA physical interactions with water, ions and proteins/enzymes in living cells.

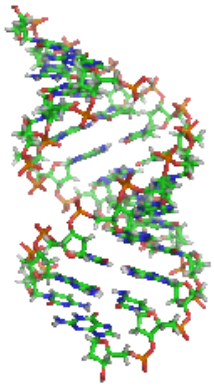
Animated DNA molecular models and hydrogen-bonding

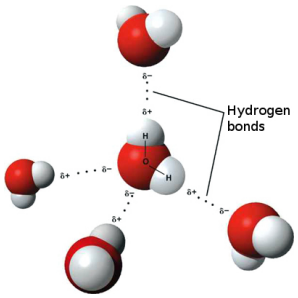
Animated molecular models allow one to visually explore the three-dimensional (3D) structure of DNA. The first DNA model is a space-filling, or CPK, model of the DNA double-helix whereas the third is an animated wire, or skeletal type, molecular model of DNA. The last two DNA molecular models in this series depict quadruplex DNA [13] that may be involved in certain cancers[11] [12] . The first CPK model in the second row is a molecular model of hydrogen bonds between water molecules in ice that are broadly similar to those found in DNA; the hydrogen bonding dynamics and proton exchange is however very different by many orders of magnitude between the two systems of fully hydrated DNA and water molecules in ice. Thus, the DNA dynamics is complex, involving nanosecond and several tens of picosecond time scales, whereas that of liquid ice is on the picosecond time scale, and that of proton exchange in ice is on the millisecond time scale; the proton exchange rates in DNA and attached proteins may vary from picosecond to nanosecond, minutes or years, depending on the exact locations of the exchanged protons in the large biopolymers. The simple harmonic oscillator 'vibration' in the third, animated image of the next gallery is only an oversimplified dynamic representation of the longitudinal vibrations of the DNA intertwined helices which were found to be anharmonic rather than harmonic as often assumed in quantum dynamic simulations of DNA.

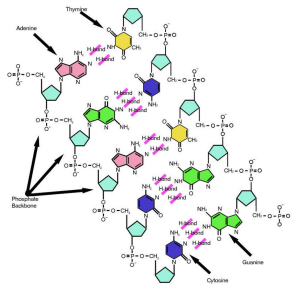


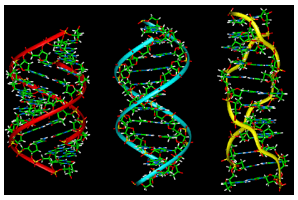


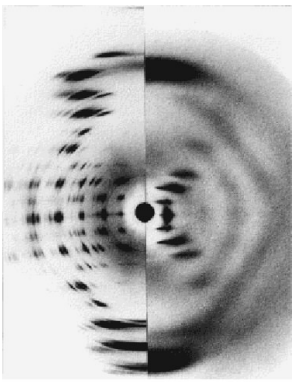


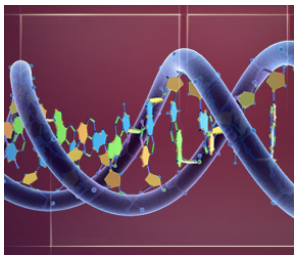


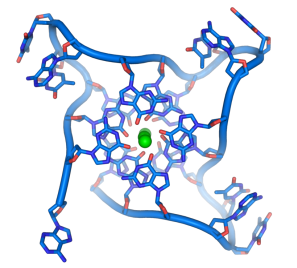

Hydrogen
bonds

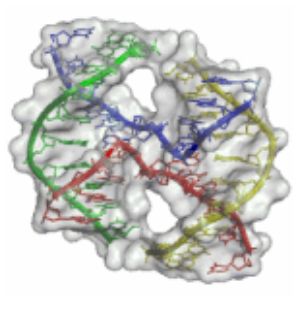





A-DNA B-DNA





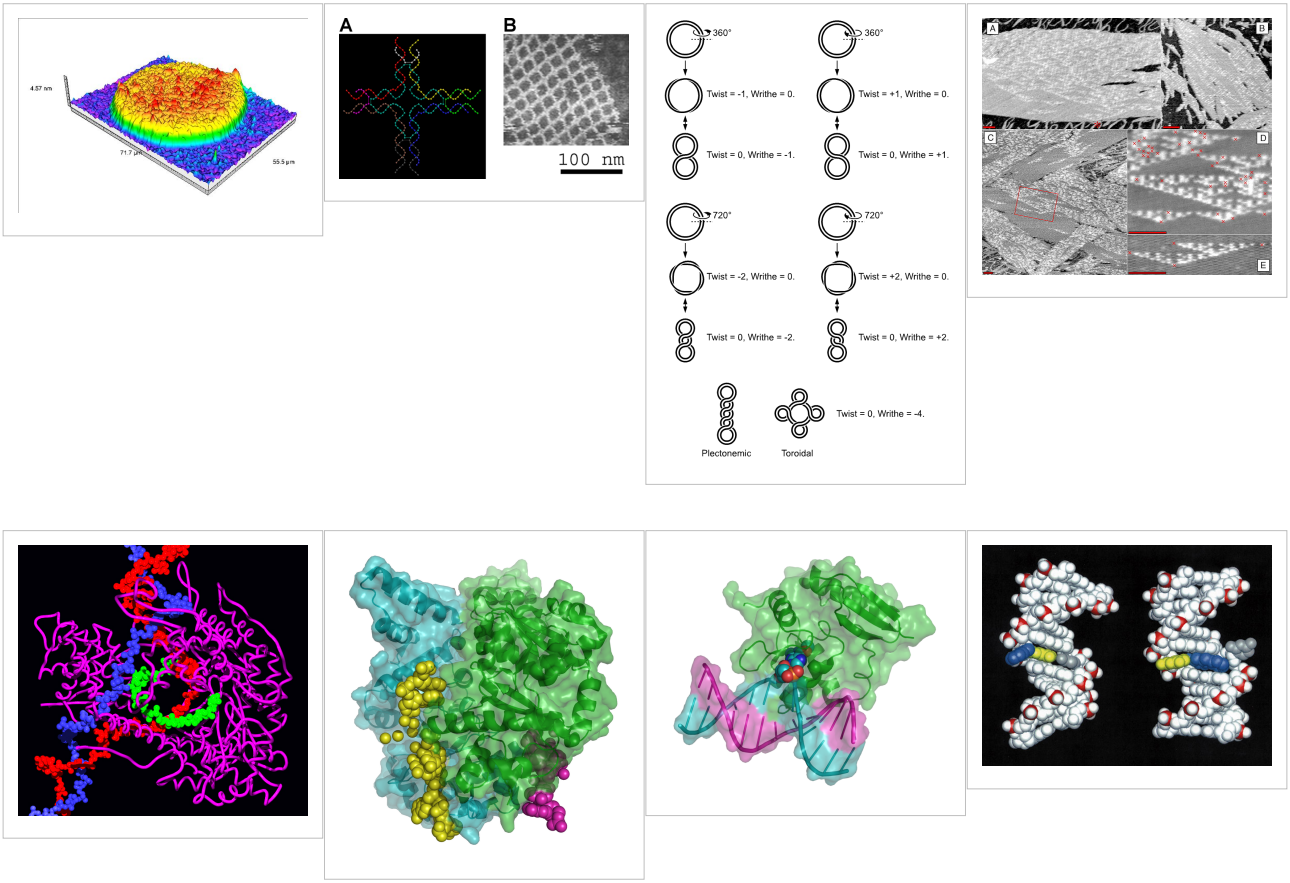


Human Genomics and Biotechnology Applications of DNA Molecular Modeling

The following two galleries of images illustrate various uses of DNA molecular modeling in Genomics and Biotechnology research applications from DNA repair to PCR and DNA nanostructures; each slide contains its own explanation and/or details. The first slide presents an overview of DNA applications, including DNA molecular models, with emphasis on Genomics and Biotechnology.

Applications of DNA molecular dynamics computations

- *First row* images present a DNA biochip and DNA nanostructures designed for DNA computing and other dynamic applications of DNA nanotechnology; last image in this row is of DNA arrays that display a representation of the Sierpinski gasket on their surfaces.
- *Second row*: the first two images show computer molecular models of RNA polymerase, followed by that of an E. coli, bacterial DNA primase template suggesting very complex dynamics at the interfaces between the enzymes and the DNA template; the fourth image illustrates in a computed molecular model the mutagenic, chemical interaction of a potent carcinogen molecule with DNA, and the last image shows the different interactions of specific fluorescence labels with DNA in human and orangoutan chromosomes.



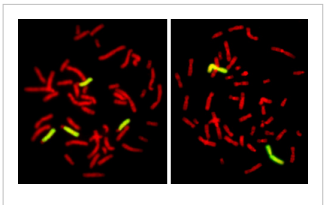
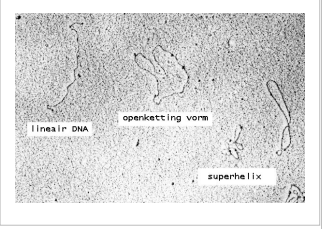
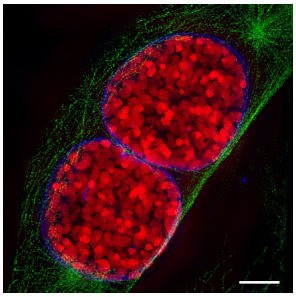
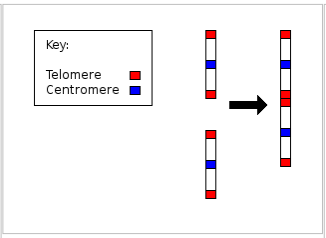


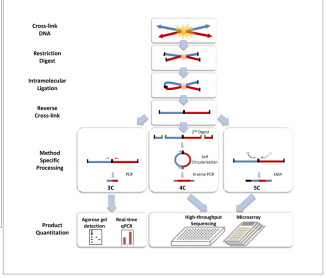
Image Gallery: *DNA Applications and Technologies at various scales in Biotechnology and Genomics research*

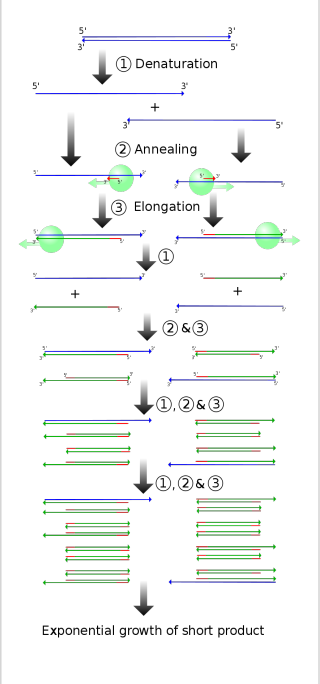
The first figure is an actual electron micrograph of a DNA fiber bundle, presumably of a single plasmid, bacterial DNA loop.

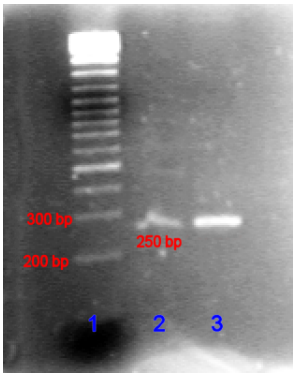


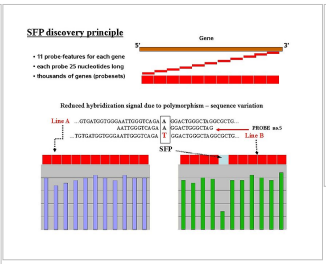


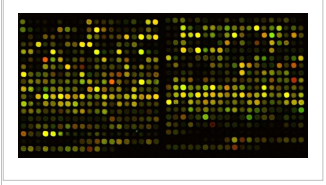


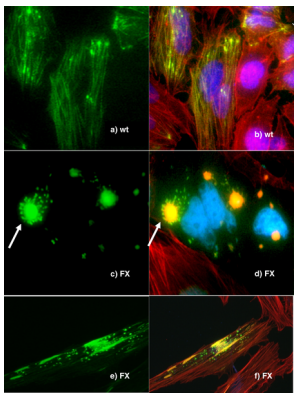


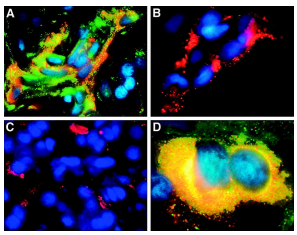


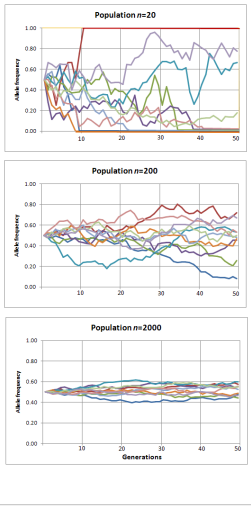


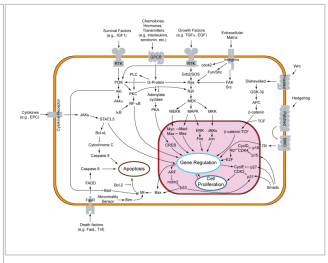


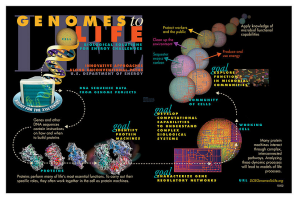










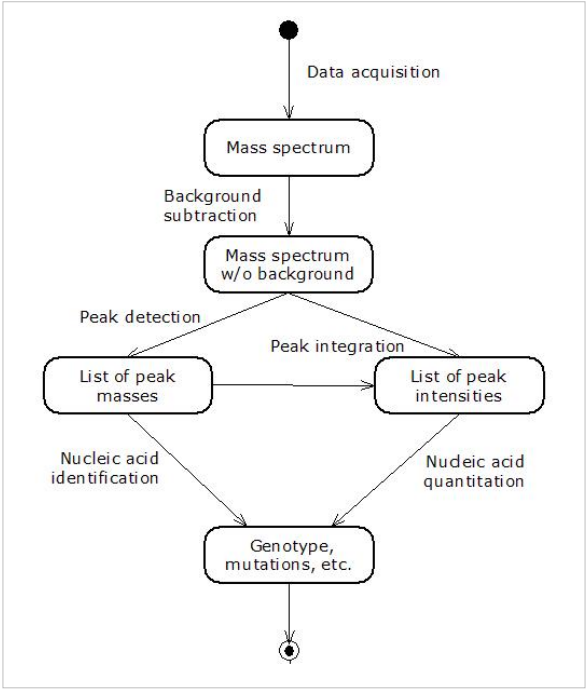


Databases for Genomics, DNA Dynamics and Sequencing

Genomic and structural databases

- CBS Genome Atlas Database [57] — contains examples of base skews.[13]
- The Z curve database of genomes — a 3-dimensional visualization and analysis tool of genomes [59][14] .
- DNA and other nucleic acids' molecular models: Coordinate files of nucleic acids molecular structure models in PDB and CIF formats [61]

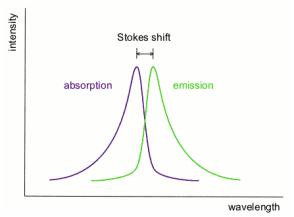
Mass spectrometry--Maldi informatics

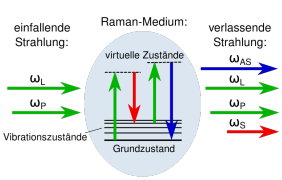


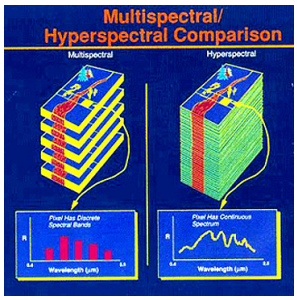
DNA Dynamics Data from Spectroscopy

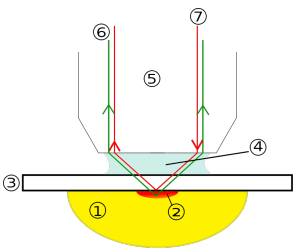
- FT-NMR[15] [16]
  - NMR Atlas--database [29]
  - mmcif downloadable coordinate files of nucleic acids in solution from 2D-FT NMR data [30]
  - NMR constraints files for NAs in PDB format [31]
- NMR microscopy[17]
- Vibrational circular dichroism (VCD)
- Microwave spectroscopy
- FT-IR
- FT-NIR[18] [19] [20]
- Spectral, Hyperspectral, and Chemical imaging)[21] [22] [23] [24] [25] [26] [27] .
- Raman spectroscopy/microscopy[28] and CARS[29] .
- Fluorescence correlation spectroscopy[30] [31] [32] [33] [34] [35] [36] [37] , Fluorescence cross-correlation spectroscopy and FRET[38] [39] [40] .
- Confocal microscopy[41]

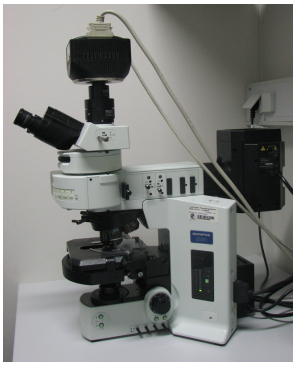
Gallery: CARS (Raman spectroscopy), Fluorescence confocal microscopy, and Hyperspectral imaging

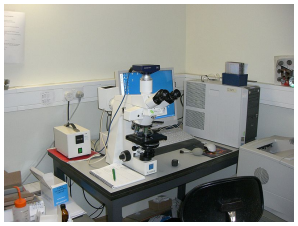


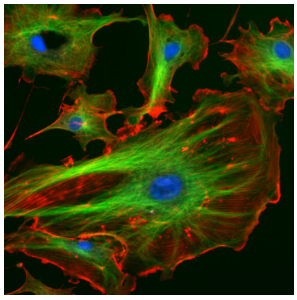


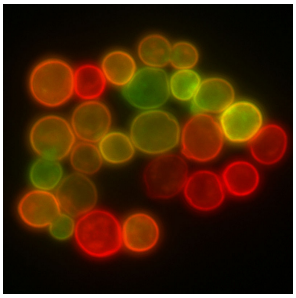


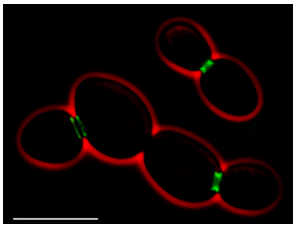


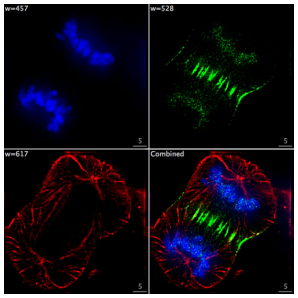


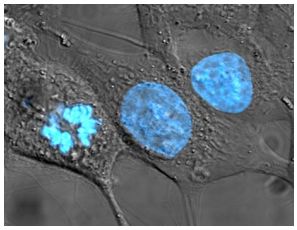


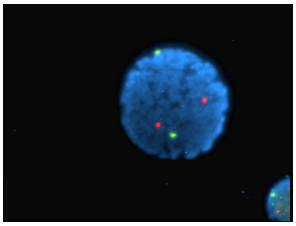


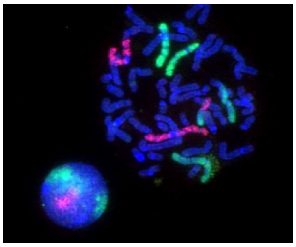


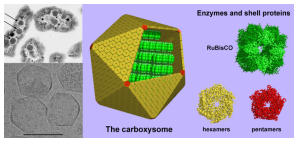












X-ray microscopy

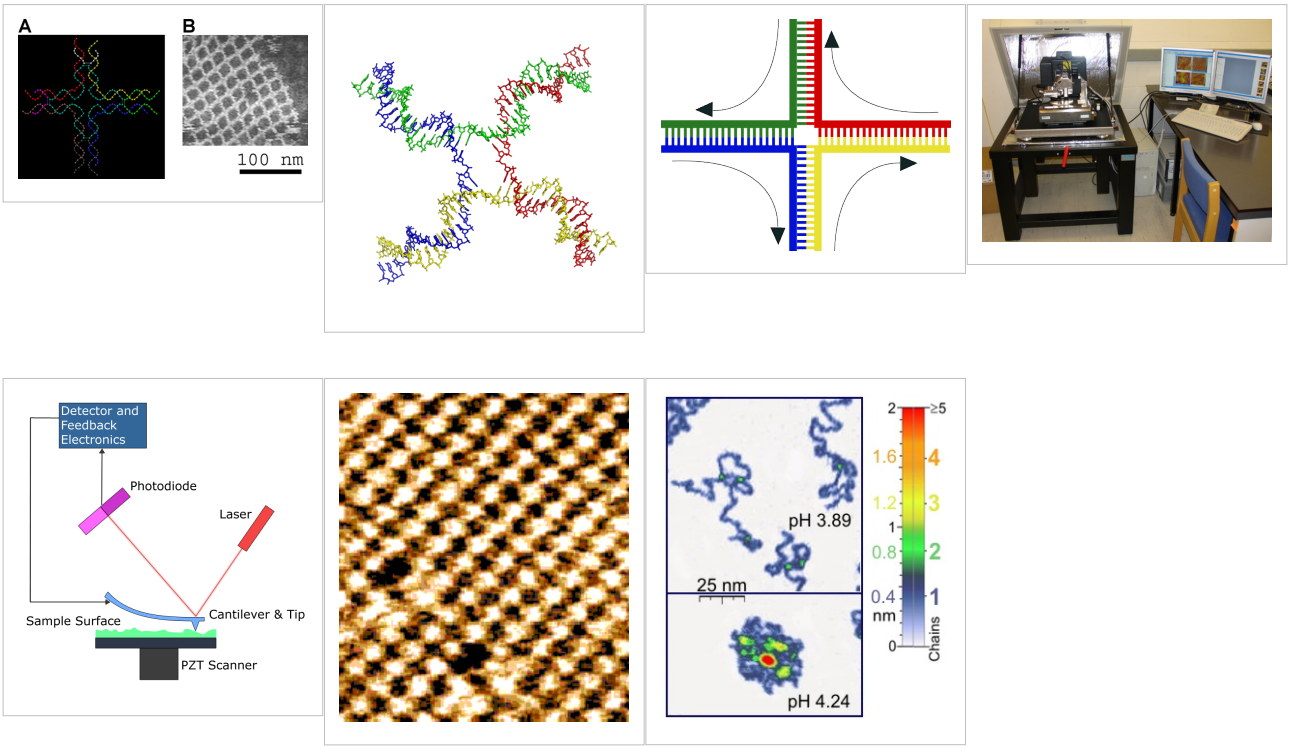
- Application of X-ray microscopy in the analysis of living hydrated cells [18]

Atomic Force Microscopy (AFM)

Two-dimensional DNA junction arrays have been visualized by Atomic Force Microscopy (AFM)[42] . Other imaging resources for AFM/Scanning probe microscopy(SPM) can be freely accessed at:

- How SPM Works [25]
- SPM Image Gallery - AFM STM SEM MFM NSOM and more. [26]

Gallery of AFM Images of DNA Nanostructures



Notes

[1] Franklin, R.E. and Gosling, R.G. recd.6 March 1953. Acta Cryst. (1953). 6, 673 The Structure of Sodium Thymonucleate Fibres I. The Influence of Water Content Acta Cryst. (1953). and 6, 678 The Structure of Sodium Thymonucleate Fibres II. The Cylindrically Symmetrical Patterson Function.

[2] Watson, J.D; Crick F.H.C. 1953a. Molecular Structure of Nucleic Acids- A Structure for Deoxyribose Nucleic Acid., Nature 171(4356):737-738.

[3] Watson, J.D; Crick F.H.C. 1953b. The Structure of DNA., Cold Spring Harbor Symposia on Quantitative Biology 18:123-131.

[4] Wilkins M.H.F., A.R. Stokes A.R. & Wilson, H.R. (1953). "http://www.nature.com/nature/dna50/wilkins.pdf|Molecular Structure of Deoxypentose Nucleic Acids" (PDF). *Nature* **171**: 738–740. doi: 10.1038/171738a0 (http://dx.doi.org/10.1038/171738a0). PMID 13054693. http://www.nature.com/nature/dna50/wilkins.pdf.

[5] Elson D, Chargaff E (1952). "On the deoxyribonucleic acid content of sea urchin gametes". *Experientia* **8** (4): 143-145.

[6] Chargaff E, Lipshitz R, Green C (1952). "Composition of the deoxypentose nucleic acids of four genera of sea-urchin". *J Biol Chem* **195** (1): 155-160. PMID 14938364.

[7] Chargaff E, Lipshitz R, Green C, Hodes ME (1951). "The composition of the deoxyribonucleic acid of salmon sperm". *J Biol Chem* **192** (1): 223-230. PMID 14917668.

[8] Chargaff E (1951). "Some recent studies on the composition and structure of nucleic acids". *J Cell Physiol Suppl* **38** (Suppl).
[9] Magasanik B, Vischer E, Doniger R, Elson D, Chargaff E (1950). "The separation and estimation of ribonucleotides in minute quantities". *J Biol Chem* **186** (1): 37-50. PMID 14778802.
[10] Chargaff E (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation". *Experientia* **6** (6): 201-209.
[11] http://www.phy.cam.ac.uk/research/bss/molbiophysics.php
[12] http://planetphysics.org/encyclopedia/TheoreticalBiophysics.html
[13] Hallin PF, David Ussery D (2004). "CBS Genome Atlas Database: A dynamic storage for bioinformatic results and DNA sequence data". *Bioinformatics* **20**: 3682-3686.
[14] Zhang CT, Zhang R, Ou HY (2003). "The Z curve database: a graphic representation of genome sequences". Bioinformatics 19 (5): 593-599. doi:10.1093/bioinformatics/btg041
[15] (http://www.jonathanpmiller.com/Karplus.html)- obtaining dihedral angles from ^{3}J coupling constants
[16] (http://www.spectroscopynow.com/FCKeditor/UserFiles/File/specNOW/HTML files/General_Karplus_Calculator.htm) Another Javascript-like NMR coupling constant to dihedral
[17] Lee, S. C. et al., (2001). One Micrometer Resolution NMR Microscopy. *J. Magn. Res.*, **150**: 207-213.
[18] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[19] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[20] Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
[21] http://www.imaging.net/chemical-imaging/Chemical imaging
[22] http://www.malvern.com/LabEng/products/sdi/bibliography/sdi_bibliography.htm E. N. Lewis, E. Lee and L. H. Kidder, Combining Imaging and Spectroscopy: Solving Problems with Near-Infrared Chemical Imaging. Microscopy Today, Volume 12, No. 6, 11/2004.
[23] D.S. Mantus and G. H. Morrison. 1991. Chemical imaging in biology and medicine using ion microscopy., *Microchimica Acta*, **104**, (1-6) January 1991, doi: 10.1007/BF01245536
[24] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[25] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann and S. S. Korban, q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[26] J. Dubois, G. Sando, E. N. Lewis, Near-Infrared Chemical Imaging, A Valuable Tool for the Pharmaceutical Industry, G.I.T. Laboratory Journal Europe, No.1-2, 2007.
[27] Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology.(June 2004).,I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin q-bio/0406047 (http://arxiv.org/abs/q-bio/0406047)
[28] Chemical Imaging Without Dyeing (http://witec.de/en/download/Raman/ImagingMicroscopy04.pdf)
[29] C.L. Evans and X.S. Xie.2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine., doi:10.1146/annurev.anchem.1.031207.112754 *Annual Review of Analytical Chemistry*, **1**: 883-909.
[30] Eigen, M., Rigler, M. Sorting single molecules: application to diagnostics and evolutionary biotechnology,(1994) *Proc. Natl. Acad. Sci. USA*, 91,5740-5747.
[31] Rigler, M. Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology,(1995) J. Biotechnol., 41,177-186.
[32] Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
[33] Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004.I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004) (http://arxiv.org/abs/q-bio/0407006)
[34] Oehlenschläger F., Schwille P. and Eigen M. (1996). Detection of HIV-1 RNA by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. USA* **93**:1281.
[35] Bagatolli, L.A., and Gratton, E. (2000). Two-photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. Biophys J., 78:290-305.

[36] Schwille, P., Haupts, U., Maiti, S., and Webb. W.(1999). Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. Biophysical Journal, 77(10):2251-2265.
[37] Near Infrared Microspectroscopy, Fluorescence Microspectroscopy,Infrared Chemical Imaging and High Resolution Nuclear Magnetic Resonance Analysis of Soybean Seeds, Somatic Embryos and Single Cells., Baianu, I.C. et al. 2004., In *Oil Extraction and Analysis.*, D. Luthria, Editor pp.241-273, AOCS Press., Champaign, IL.
[38] FRET description (http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/fret.htm)
[39] doi:10.1016/S0959-440X(00)00190-1 (http://dx.doi.org/10.1016/S0959-440X(00)00190-1)Recent advances in FRET: distance determination in protein–DNA complexes. *Current Opinion in Structural Biology* **2001**, *11(2)*, 201-207
[40] http://www.fretimaging.org/mcnamaraintro.html FRET imaging introduction
[41] Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, Proc. Natl. Acad. Sci. USA 91:5740.
[42] Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).

References

- Sir Lawrence Bragg, FRS. *The Crystalline State, A General survey.* London: G. Bells and Sons, Ltd., vols. 1 and 2., 1966., 2024 pages.
- F. Bessel, *Untersuchung des Theils der planetarischen Störungen*, Berlin Abhandlungen (1824), article 14.
- Cantor, C. R. and Schimmel, P.R. *Biophysical Chemistry, Parts I and II.*, San Franscisco: W.H. Freeman and Co. 1980. 1,800 pages.
- Eigen, M., and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology, *Proc. Natl. Acad. Sci. USA* **91**:5740.
- Raghavachari, R., Editor. 2001. *Near-Infrared Applications in Biotechnology*, Marcel-Dekker, New York, NY.
- Rigler R. and Widengren J. (1990). Ultrasensitive detection of single molecules by fluorescence correlation spectroscopy, *BioScience* (Ed. Klinge & Owman) p.180.
- *Applications of Novel Techniques to Health Foods, Medical and Agricultural Biotechnology*.(June 2004) I. C. Baianu, P. R. Lozano, V. I. Prisecaru and H. C. Lin., q-bio/0406047.
- Single Cancer Cell Detection by Near Infrared Microspectroscopy, Infrared Chemical Imaging and Fluorescence Microspectroscopy.2004. I. C. Baianu, D. Costescu, N. E. Hofmann, S. S. Korban and et al., q-bio/0407006 (July 2004).
- Voet, D. and J.G. Voet. *Biochemistry*, 2nd Edn., New York, Toronto, Singapore: John Wiley & Sons, Inc., 1995, ISBN 0-471-58651-X., 1361 pages.
- Watson, G. N. *A Treatise on the Theory of Bessel Functions.*, (1995) Cambridge University Press. ISBN 0-521-48391-3.
- Watson, James D. and Francis H.C. Crick. A structure for Deoxyribose Nucleic Acid (http://www.nature.com/nature/dna50/watsoncrick.pdf) (PDF). *Nature* 171, 737–738, 25 April 1953.
- Watson, James D. *Molecular Biology of the Gene.* New York and Amsterdam: W.A. Benjamin, Inc. 1965., 494 pages.
- Wentworth, W.E. *Physical Chemistry. A short course.*, Malden (Mass.): Blackwell Science, Inc. 2000.

- Herbert R. Wilson, FRS. *Diffraction of X-rays by proteins, Nucleic Acids and Viruses.*, London: Edward Arnold (Publishers) Ltd. 1966.
- Kurt Wuthrich. *NMR of Proteins and Nucleic Acids.*, New York, Brisbane,Chicester, Toronto, Singapore: J. Wiley & Sons. 1986., 292 pages.
- Robinson, Bruche H.; Seeman, Nadrian C. (August 1987). "The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device". *Protein Engineering* **1** (4): 295–300. ISSN 0269-2139 (http://worldcat.org/issn/0269-2139). Link (http://peds.oxfordjournals.org/cgi/content/abstract/1/4/295)
- Rothemund, Paul W. K.; Ekani-Nkodo, Axel; Papadakis, Nick; Kumar, Ashish; Fygenson, Deborah Kuchnir & Winfree, Erik (22 December 2004). "Design and Characterization of Programmable DNA Nanotubes". *Journal of the American Chemical Society* **126** (50): 16344–16352. doi: 10.1021/ja044319l (http://dx.doi.org/10.1021/ja044319l). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Keren, K.; Kinneret Keren, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, Erez Braun (November 2003). "http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380|DNA-Templated Carbon Nanotube Field-Effect Transistor". *Science* **302** (6549): 1380–1382. doi: 10.1126/science.1091022 (http://dx.doi.org/10.1126/science.1091022). ISSN 1095-9203 (http://worldcat.org/issn/1095-9203). http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380.
- Zheng, Jiwen; Constantinou, Pamela E.; Micheel, Christine; Alivisatos, A. Paul; Kiehl, Richard A. & Seeman Nadrian C. (2006). "2D Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs". *Nano Letters* **6**: 1502–1504. doi: 10.1021/nl060994c (http://dx.doi.org/10.1021/nl060994c). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Cohen, Justin D.; Sadowski, John P.; Dervan, Peter B. (2007). "Addressing Single Molecules on DNA Nanostructures". *Angewandte Chemie* **46** (42): 7956–7959. doi: 10.1002/anie.200702767 (http://dx.doi.org/10.1002/anie.200702767). ISSN 0570-0833 (http://worldcat.org/issn/0570-0833).
- Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Constantinou, Pamela E.; Wang, Tong; Kopatsch, Jens; Israel, Lisa B.; Zhang, Xiaoping; Ding, Baoquan; Sherman, William B.; Wang, Xing; Zheng, Jianping; Sha, Ruojie & Seeman, Nadrian C. (2006). "Double cohesion in structural DNA nanotechnology". *Organic and Biomolecular Chemistry* **4**: 3414–3419. doi: 10.1039/b605212f (http://dx.doi.org/10.1039/b605212f).

See also

- DNA
- Molecular modeling of DNA
- Genomics
- Signal transduction
- Transcriptomics
- Interactomics
- Biotechnology
- Molecular graphics
- Quantum computing
- MAYA-II
- DNA computing
- DNA structure
- Molecular structure
- Molecular dynamics
- Molecular topology
- DNA topology
- DNA, the Genome and Interactome
- Molecular structure
- Molecular geometry fluctuations
- Molecular interactions
- Molecular topology
- Hydrogen bonding
- Hydrophobic interactions
- DNA dynamics and conformations
- DNA Conformational isomerism
- 2D-FT NMRI and Spectroscopy
- Paracrystalline lattices/Paracrystals
- NMR Spectroscopy
- VCD or Vibrational circular dichroism
- Microwave spectroscopy
- Two-dimensional IR spectroscopy
- FRET and FCS- Fluorescence correlation spectroscopy
- Fluorescence cross-correlation spectroscopy (FCCS)
- Spectral imaging
- Hyperspectral imaging
- Chemical imaging
- NMR microscopy
- X-ray scattering
- Neutron scattering
- Crystallography
- Crystal lattices
- Molecular geometry
- Nanostructure
- DNA nanotechnology
- Imaging
- Sirius visualization software

- Atomic force microscopy
- X-ray microscopy
- Liquid crystals
- Glasses
- QMC@Home
- Sir Lawrence Bragg, FRS
- Sir John Randall
- Francis Crick
- Manfred Eigen
- Felix Bloch
- Paul Lauterbur
- Maurice Wilkins
- Herbert Wilson, FRS
- Alex Stokes

External links

- DNAlive: a web interface to compute DNA physical properties (http://mmb.pcb.ub.es/DNAlive). Also allows cross-linking of the results with the UCSC Genome browser and DNA dynamics.
- Application of X-ray microscopy in analysis of living hydrated cells (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=12379938)
- DiProDB: Dinucleotide Property Database (http://diprodb.fli-leibniz.de). The database is designed to collect and analyse thermodynamic, structural and other dinucleotide properties.
- DNA the Double Helix Game (http://nobelprize.org/educational_games/medicine/dna_double_helix/) From the official Nobel Prize web site
- MDDNA: Structural Bioinformatics of DNA (http://humphry.chem.wesleyan.edu:8080/MDDNA/)
- Double Helix 1953–2003 (http://www.ncbe.reading.ac.uk/DNA50/) National Centre for Biotechnology Education
- DNA under electron microscope (http://www.fidelitysystems.com/Unlinked_DNA.html)
- Further details of mathematical and molecular analysis of DNA structure based on X-ray data (http://planetphysics.org/encyclopedia/BesselFunctionsApplicationsToDiffractionByHelicalStructures.html)
- Bessel functions corresponding to Fourier transforms of atomic or molecular helices. (http://planetphysics.org/?op=getobj&from=objects&name=BesselFunctionsAndTheirApplicationsToDiffractionByHelicalStructures)
- Characterization in nanotechnology some pdfs (http://nanocharacterization.sitesled.com/)
- An overview of STM/AFM/SNOM principles with educative videos (http://www.ntmdt.ru/SPM-Techniques/Principles/)
- SPM Image Gallery - AFM STM SEM MFM NSOM and More (http://www.rhk-tech.com/results/showcase.php)
- How SPM Works (http://www.parkafm.com/New_html/resources/01general.php)

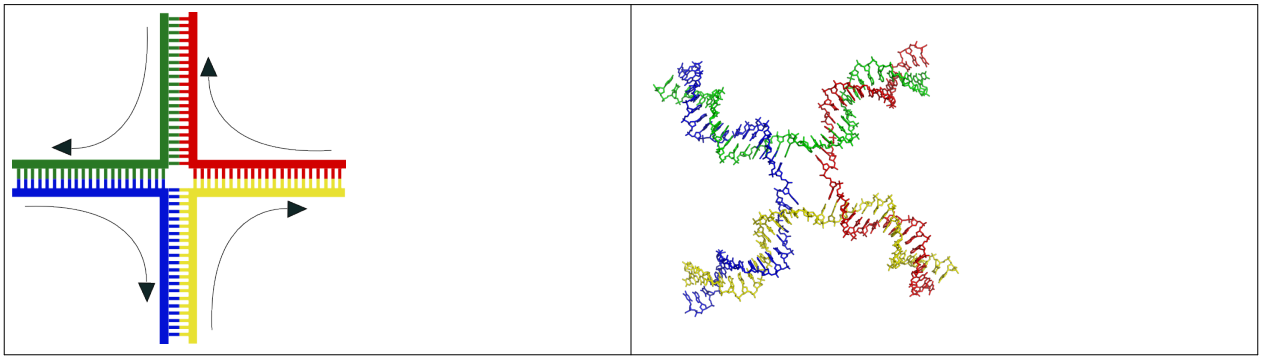
- U.S. National DNA Day (http://www.genome.gov/10506367) — watch videos and participate in real-time discussions with scientists.
- The Secret Life of DNA - DNA Music compositions (http://www.tjmitchell.com/stuart/dna.html)
- Ascalaph DNA (http://www.agilemolecule.com/Ascalaph/Ascalaph_DNA.html) — Commercial software for DNA modeling

DNA nanotechnology

Part of a series of articles on **Molecular self-assembly**
Self-assembled monolayer Supramolecular assembly DNA nanotechnology
See also Nanotechnology

DNA nanotechnology is a subfield of nanotechnology which seeks to use the unique molecular recognition properties of DNA and other nucleic acids to create novel, controllable structures out of DNA. The DNA is thus used as a structural material rather than as a carrier of genetic information, making it an example of bionanotechnology. This has possible applications in molecular self-assembly and in DNA computing.

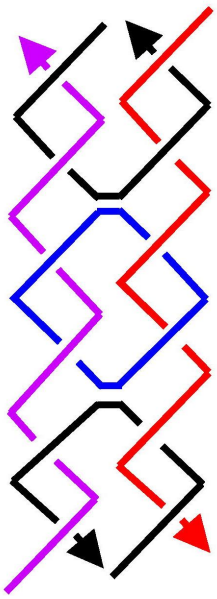
Introduction: DNA crossover molecules



Structure of the 4-arm junction.
Left: A schematic. **Right**: A more realistic model.[1]
Each of the four separate DNA single strands are shown in different colors.

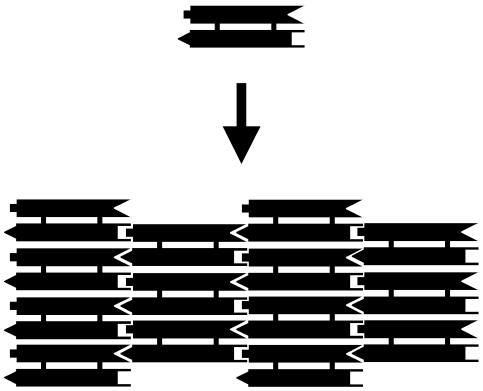
DNA nanotechnology makes use of branched DNA structures to create DNA complexes with useful properties. DNA is normally a linear molecule, in that its axis is unbranched. However, DNA molecules containing junctions can also be made. For example, a four-arm junction can be made using four individual DNA strands which are complementary to each other in the correct pattern. Due to Watson-Crick base pairing, only portions of the strands which are complementary to each other will attach to each other to form duplex DNA. This four-arm junction is an immobile form of a Holliday junction.

Junctions can be used in more complex molecules. The most important of these is the "double-crossover" or DX motif. Here, two DNA duplexes lie next to each other, and share two junction points where strands cross from one duplex into the other. This molecule has the advantage that the junction points are now constrained to a single orientation as opposed to being flexible as in the four-arm junction. This makes the DX motif suitible as a structural building block for larger DNA complexes.[3]



A double-crossover (DX) molecule. This molecule consists of five DNA single strands which form two double-helical domains, on the left and the right in this image. There are two crossover points where the strands cross from one domain into the other. Image from Mao, 2004. [2]

Tile-based arrays



Assembly of a DX array. Each bar represents a double-helical domain of DNA, with the shapes representing comlimentary sticky ends. The DX molecule at top will combine into the two-dimensional DNA array shown at bottom. Image from Mao, 2004. [2]

DX arrays

DX, Double Crossover, molecules can be equipped with sticky ends in order to combine them into a two-dimenstional periodic lattice. Each DX molecule has four termini, one at each end of the two double-helical domains, and these can be equipped with sticky ends that program them to combine into a specific pattern. More than one type of DX can be used which can be made to arrange in rows or any other tessellated pattern. They thus form extended flat sheets which are essentially two-dimensional crystals of DNA.[4]

DNA nanotubes

In addition to flat sheets, DX arrays have been made to form hollow tubes of 4-20 nm diameter. These DNA nanotubes are somewhat similar in size and shape to carbon nanotubes, but the carbon nanotubes are stronger and better conductors, whereas the DNA nanotubes are more easily modified and connected to other structures.[5]

Other tile arrays

Two-dimensional arrays have been made out of other motifs as well, including the Holliday junction rhombus array as well as various DX-based arrays in the shapes of triangles and hexagons.[6] Another motif, the six-helix bundle, has the ability to form three-dimensional DNA arrays as well.[7]

DNA origami

As an alternative to the tile-based approach, two-dimensional DNA structures can be made from a single, long DNA strand of arbitrary sequence which is folded into the desired shape by using shorter, "staple" strands. This allows the creation of two-dimensional shapes at the nanoscale using DNA. Demonstrated designs have included the smiley face and a coarse map of North America. DNA origami was the cover story of *Nature* on March 15, 2006.[8]

DNA polyhedra

A number of three-dimensional DNA molecules have been made which have the connectivity of a polyhedron such as an octahedron or cube. In other words, the DNA duplexes trace the edges of a polyhedron with a DNA junction at each vertex. The earliest demonstrations of DNA polyhedra involved multiple ligations and solid-phase synthesis steps to create catenated polyhedra. More recently, there have been demonstrations of a DNA truncated octahedron made from a long single strand designed to fold into the correct conformation, as well as a tetrahedron which can be produced from four DNA strands in a single step.[9]

DNA nanomechanical devices

DNA complexes have been made which change their conformation upon some stimulus. These are intended to have applications in nanorobotics. One of the first such devices, called "molecular tweezers," changes from an open to a closed state based upon the presence of control strands.

DNA machines have also been made which show a twisting motion. One of these makes use of the transition between the B-DNA and Z-DNA forms to respond to a change in buffer conditions. Another relies on the presence of control strands to switch from a paranemic-crossover (PX) conformation to a double-junction (JX2) conformation.[10]

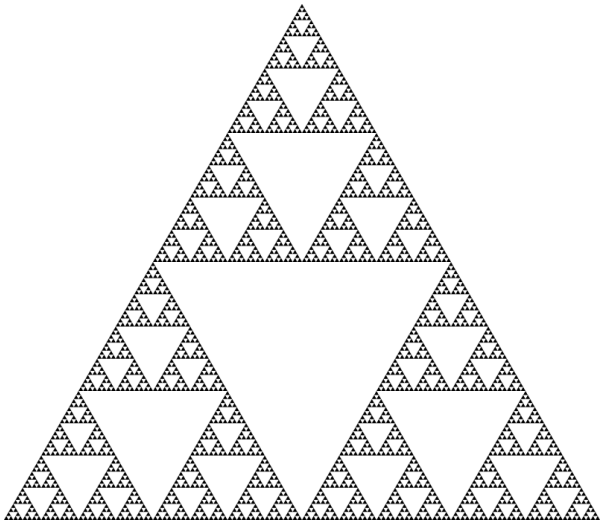
Stem Loop Controllers

A design called a *stem loop*, consisting of a single strand of DNA which has a loop at an end, are a dynamic structure that opens and closes when a piece of DNA bonds to the loop part. This effect has been exploited to create several logic gates. [11] [12] These logic gates have been used to create the computers MAYA I and MAYA II which can play tick-tac-toe to some extent.[13]

Applications

Algorithmic self-assembly

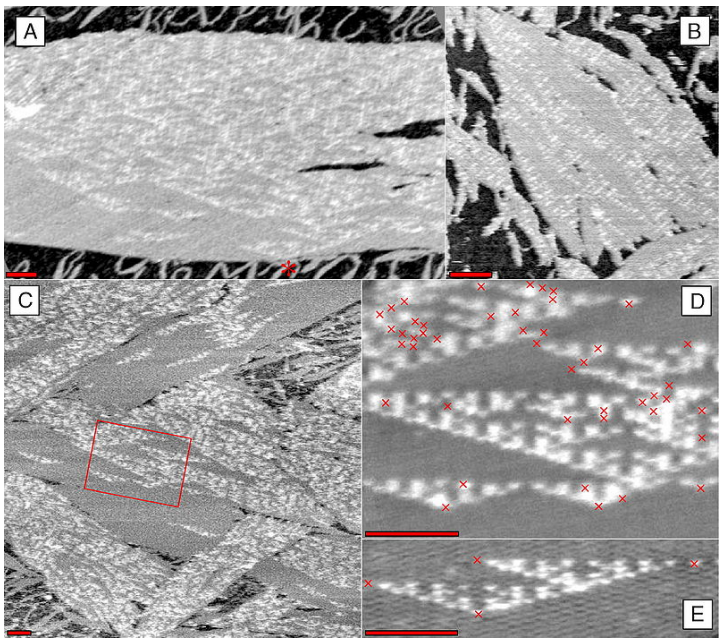
DNA nanotechnology has been applied to the related field of DNA computing. A DX array has been demonstrated whose assembly encodes an XOR operation, which allows the DNA array to implement a cellular automaton which generates a fractal called the Sierpinski gasket. This shows that computation can be incorporated into the assembly of DNA arrays, increasing its scope beyond simple periodic arrays.



The Sierpinski gasket.

Note that DNA computing overlaps with, but is distinct from, DNA nanotechnology. The latter uses the specificity of Watson-Crick basepairing to make novel structures out of DNA. These structures can be used for DNA computing, but they do not have to be. Additionally, DNA computing can be done without using the types of molecules made possible by DNA Nanotechnology.[15]

Nanoarchitecture



DNA arrays that display a representation of the Sierpinski gasket on their surfaces. Click the image for further details. Image from Rothemund *et al.*, 2004. [14]

The idea of using DNA arrays to template the assembly of other functional molecules has been around for a while, but only recently has progress been made in reducing these kinds of schemes to practice. In 2006, researchers covalently attached gold nanoparticles to a DX-based tile and showed that self-assembly of the DNA structures also assembled the nanoparticles hosted on them. A non-covalent hosting scheme was shown in 2007, using Dervan polyamides on a DX array to arrange streptavidin proteins on specific kinds of tiles on the DNA array.[16] Previously in 2006 LaBean demonstrated the letters "D" "N" and "A" created on a 4x4 DX array using streptavidin. [17]

DNA has also been used to assemble a single walled carbon nanotube Field-effect transistor.[18]

See also

- Mechanical properties of DNA

External links

- Chengde Mao page at Purdue University [19]
- John Reif lab at Duke University [20]
- Nadrian Seeman lab at NYU [21]
- William M. Shih lab at Harvard Medical School [22]
- Andrew Turberfield lab at Oxford University [23]
- Erik Winfree lab at Caltech [24]
- Hao Yan lab at Arizona State University [25]
- Bernard Yurke formerly at Bell Labs [26] now at Boise State University [27]
- Thom LaBean at Duke University [28]
- Software for 3D DNA design, modeling and/or simulation:
  - Ascalaph Designer [29]
  - caDNAno [30]
  - GIDEON [31]
  - NanoEngineer-1 [32]
- International Society for Nanoscale Science, Computation and Engineering [33]

References

Note: Click on the doi to access the text of the referenced article.

[1] Created from PDB 1M6G (http://www.rcsb.org/pdb/explore/explore.do?structureId=1M6G)
[2] http://dx.doi.org/10.1371/journal.pbio.0020431

- Seeman, Nadrian C. (1 November 1999). "DNA Engineering and its Application to Nanotechnology". *Trends in Biotechnology* **17** (11): 437–443. doi: 10.1016/S0167-7799(99)01360-8 (http://dx.doi.org/10.1016/S0167-7799(99)01360-8). ISSN 0167-7799 (http://worldcat.org/issn/0167-7799).
- Seeman, Nadrian C. (January 2001). "DNA Nicks and Nodes and Nanotechnology". *Nano Letters* **1** (1): 22–26. doi: 10.1021/nl000182v (http://dx.doi.org/10.1021/nl000182v). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Mao, Chengde (December 2004). "The Emergence of Complexity: Lessons from DNA". *PLoS Biology* **2** (12): 2036–2038. doi: 10.1371/journal.pbio.0020431 (http://dx.doi.org/10.1371/journal.pbio.0020431). ISSN 1544-9173 (http://worldcat.org/issn/1544-9173).
- Kumara, Mudalige T. (July 2008). "Assembly pathway analysis of DNA nanostructures and the construction of parallel motifs". *Nano Letters* **8** (7): 1971–1977. doi: 10.1021/nl800907y (http://dx.doi.org/10.1021/nl800907y). ISSN .
- Winfree, Erik; Liu, Furong; Wenzler, Lisa A. & Seeman, Nadrian C. (6 August 1998). "Design and self-assembly of two-dimensional DNA crystals". *Nature* **394**: 529–544. doi: 10.1038/28998 (http://dx.doi.org/10.1038/28998). ISSN 0028-0836 (http://worldcat.org/issn/0028-0836).
- Liu, Furong; Sha, Ruojie & Seeman, Nadrian C. (10 February 1999). "Modifying the Surface Features of Two-Dimensional DNA Crystals". *Journal of the American Chemical Society* **121** (5): 917–922. doi: 10.1021/ja982824a (http://dx.doi.org/10.1021/ja982824a). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Rothemund, Paul W. K.; Ekani-Nkodo, Axel; Papadakis, Nick; Kumar, Ashish; Fygenson, Deborah Kuchnir & Winfree, Erik (22 December 2004). "Design and Characterization of Programmable DNA Nanotubes". *Journal of the American Chemical Society* **126** (50): 16344–16352. doi: 10.1021/ja044319l (http://dx.doi.org/10.1021/ja044319l). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Mao, Chengde; Sun, Weiqiong & Seeman, Nadrian C. (16 June 1999). "Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy". *Journal of the American Chemical Society* **121** (23): 5437–5443. doi: 10.1021/ja9900398 (http://dx.doi.org/10.1021/ja9900398). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Constantinou, Pamela E.; Wang, Tong; Kopatsch, Jens; Israel, Lisa B.; Zhang, Xiaoping; Ding, Baoquan; Sherman, William B.; Wang, Xing; Zheng, Jianping; Sha, Ruojie & Seeman, Nadrian C. (2006). "Double cohesion in structural DNA nanotechnology". *Organic and Biomolecular Chemistry* **4**: 3414–3419. doi: 10.1039/b605212f (http://dx.doi.org/10.1039/b605212f).
- Mathieu, Frederick; Liao, Shiping; Kopatsch, Jens; Wang, Tong; Mao, Chengde & Seeman, Nadrian C. (April 2005). "Six-Helix Bundles Designed from DNA". *Nano Letters* **5** (4): 661–665. doi: 10.1021/nl050084f (http://dx.doi.org/10.1021/nl050084f). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Rothemund, Paul W. K. (2006). "Folding DNA to create nanoscale shapes and patterns". *Nature* **440**: 297–302. doi: 10.1038/nature04586 (http://dx.doi.org/10.1038/nature04586). ISSN 0028-0836 (http://worldcat.org/issn/0028-0836).
- Zhang, Yuwen; Seeman, Nadrian C. (1994). "Construction of a DNA-truncated octahedron". *Journal of the American Chemical Society* **116** (5): 1661–1669. doi: 10.1021/ja00084a006 (http://dx.doi.org/10.1021/ja00084a006). ISSN 0002-7863 (http://worldcat.org/issn/0002-7863).
- Shih, William M.; Quispe, Joel D.; Joyce, Gerald F. (12 February 2004). "A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron". *Nature* **427**: 618–621. doi: 10.1038/nature02307 (http://dx.doi.org/10.1038/nature02307). ISSN 0028-0836 (http://worldcat.org/issn/0028-0836).
- Goodman, R.P.; Schaap, I.A.T.; Tardin, C.F.; Erben, C.M.; Berry, R.M.; Schmidt, C.F.; Turberfield, A.J. (9 December 2005). "Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication". *Science* **310** (5754): 1661–1665. doi: 10.1126/science.1120367 (http://dx.doi.org/10.1126/science.1120367). ISSN 0036-8075 (http://worldcat.org/issn/0036-8075).
- Yurke, Bernard; Turberfield, Andrew J.; Mills, Allen P., Jr; Simmel, Friedrich C. & Neumann, Jennifer L. (10 August 2000). "A DNA-fuelled molecular machine made of DNA". *Nature* **406**: 605–609. doi: 10.1038/35020524 (http://dx.doi.org/10.1038/35020524). ISSN 0028-0836 (http://worldcat.org/issn/

0028-0836).

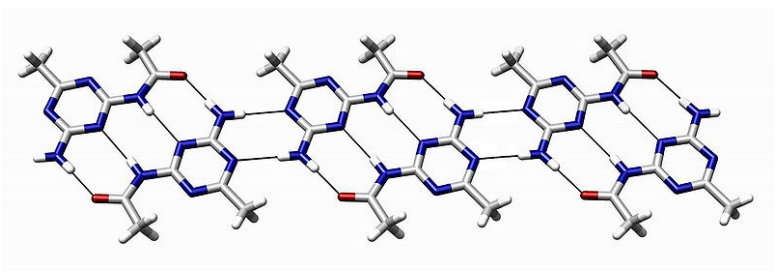
- Mao, Chengde; Sun, Weiqiong; Shen, Zhiyong & Seeman, Nadrian C. (14 January 1999). "A DNA Nanomechanical Device Based on the B-Z Transition". *Nature* **397**: 144–146. doi: 10.1038/16437 (http://dx.doi.org/10.1038/16437). ISSN .
- Yan, Hao; Zhang, Xiaoping; Shen, Zhiyong & Seeman, Nadrian C. (3 January 2002). "A robust DNA mechanical device controlled by hybridization topology". *Nature* **415**: 62–65. doi: 10.1038/415062a (http://dx.doi.org/10.1038/415062a). ISSN .

[11] DNA Logic Gates (https://digamma.cs.unm.edu/wiki/bin/view/McogPublicWeb/MolecularLogicGates)
[12] (http://www.duke.edu/~jme17/Joshua_E._Mendoza-Elias/Research_Ideas.html)
[13] MAYA II (https://digamma.cs.unm.edu/wiki/bin/view/McogPublicWeb/MolecularAutomataMAYAII)
[14] http://dx.doi.org/10.1371/journal.pbio.0020424

- Rothemund, Paul W. K.; Papadakis, Nick & Winfree, Erik (December 2004). "Algorithmic Self-Assembly of DNA Sierpinski Triangles". *PLoS Biology* **2** (12): 2041–2053. doi: 10.1371/journal.pbio.0020424 (http://dx.doi.org/10.1371/journal.pbio.0020424). ISSN 1544-9173 (http://worldcat.org/issn/1544-9173).
- Robinson, Bruche H.; Seeman, Nadrian C. (August 1987). "The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device". *Protein Engineering* **1** (4): 295–300. ISSN 0269-2139 (http://worldcat.org/issn/0269-2139). Link (http://peds.oxfordjournals.org/cgi/content/abstract/1/4/295)
- Zheng, Jiwen; Constantinou, Pamela E.; Micheel, Christine; Alivisatos, A. Paul; Kiehl, Richard A. & Seeman Nadrian C. (2006). "2D Nanoparticle Arrays Show the Organizational Power of Robust DNA Motifs". *Nano Letters* **6**: 1502–1504. doi: 10.1021/nl060994c (http://dx.doi.org/10.1021/nl060994c). ISSN 1530-6984 (http://worldcat.org/issn/1530-6984).
- Cohen, Justin D.; Sadowski, John P.; Dervan, Peter B. (2007). "Addressing Single Molecules on DNA Nanostructures". *Angewandte Chemie* **46** (42): 7956–7959. doi: 10.1002/anie.200702767 (http://dx.doi.org/10.1002/anie.200702767). ISSN 0570-0833 (http://worldcat.org/issn/0570-0833).

[17] Park, Sung Ha; Sung Ha Park, Constantin Pistol, Sang Jung Ahn, John H. Reif, Alvin R. Lebeck, Chris Dwyer, Thomas H. LaBean (October 2006). "http://www3.interscience.wiley.com/journal/113390879/abstract|Finite-Size, Fully Addressable DNA Tile Lattices Formed by Hierarchical Assembly Procedures". *Angewandte Chemie* **118** (40): 749–753. doi: 10.1002/ange.200690141 (http://dx.doi.org/10.1002/ange.200690141). ISSN 1521-3757 (http://worldcat.org/issn/1521-3757). http://www3.interscience.wiley.com/journal/113390879/abstract.
[18] Keren, K.; Kinneret Keren, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, Erez Braun (November 2003). "http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380|DNA-Templated Carbon Nanotube Field-Effect Transistor". *Science* **302** (6549): 1380–1382. doi: 10.1126/science.1091022 (http://dx.doi.org/10.1126/science.1091022). ISSN 1095-9203 (http://worldcat.org/issn/1095-9203). http://www.sciencemag.org/cgi/content/abstract/sci;302/5649/1380.
[19] http://www.chem.purdue.edu/people/faculty/faculty.asp?itemID=46
[20] http://www.cs.duke.edu/~reif/BMC/Reif.BMCproject.html
[21] http://seemanlab4.chem.nyu.edu/
[22] http://research2.dfci.harvard.edu/shih/SHIH_LAB/Home.html
[23] http://www.physics.ox.ac.uk/cm/people/turberfield.htm
[24] http://dna.caltech.edu/
[25] http://chemistry.asu.edu/faculty/hao_yan.asp
[26] http://www.bell-labs.com/org/physicalsciences/profiles/yurke.html
[27] http://coen.boisestate.edu/departments/faculty.asp?ID=134
[28] http://www.cs.duke.edu/~thl/
[29] http://www.agilemolecule.com/Ascalaph/Ascalaph_Designer.html
[30] http://cadnano.org
[31] http://www.subirac.com
[32] http://www.nanoengineer-1.net
[33] http://www.isnsce.org/

Molecular self-assembly



An example of a molecular self-assembly through hydrogen bonds reported by Meijer and coworkers.[1]

Molecular self-assembly is the process by which molecules adopt a defined arrangement without guidance or management from an outside source. There are two types of self-assembly, **intra**molecular self-assembly and **inter**molecular self-assembly. Most often the term molecular self-assembly refers to intermolecular self-assembly, while the intramolecular analog is more commonly called folding.

Supramolecular Systems

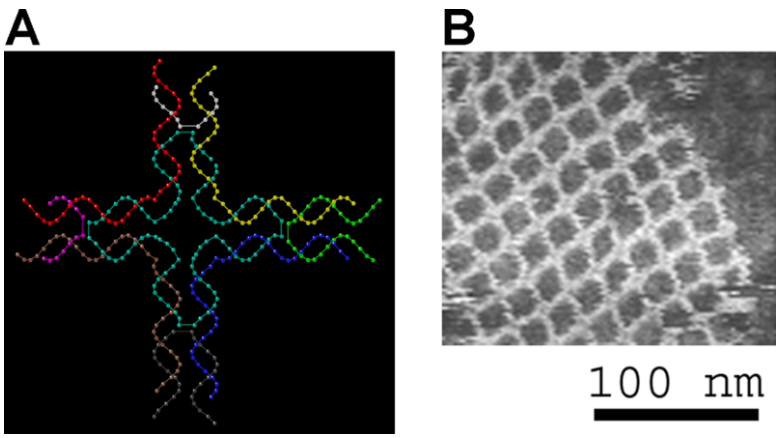
Molecular self-assembly is a key concept in supramolecular chemistry[2] [3] [4] since assembly of the molecules is directed through noncovalent interactions (e.g., hydrogen bonding, metal coordination, hydrophobic forces, van der Waals forces, π-π interactions, and/or electrostatic) as well as electromagnetic interactions. Common examples include the formation of micelles, vesicles, liquid crystal phases, and Langmuir monolayers by surfactant molecules.[5] Further examples of supramolecular assemblies demonstrate that a variety of different shapes and sizes can be obtained using molecular self-assembly.

Molecular self-assembly has allowed the construction of challenging molecular topologies. An example are Borromean rings, interlocking rings wherein removal of one ring unlocks each of the other rings. DNA has been used to prepare a molecular analog of Borromean rings.[6] More recently, a similar structure has been prepared using non-biological building blocks.[7]

Biological Systems

Molecular self-assembly is crucial to the function of cells. It is exhibited in the self-assembly of lipids to form the membrane, the formation of double helical DNA through hydrogen bonding of the individual strands, and the assembly of proteins to form quaternary structures. Molecular self-assembly of incorrectly folded proteins into insoluble amyloid fibers is responsible for infectious prion-related neurodegenerative diseases.

Nanotechnology



The DNA structure at left (schematic shown) will self-assemble into the structure visualized by atomic force microscopy at right. Image from Strong.[8]

Molecular self-assembly is an important aspect of bottom-up approaches to nanotechnology. Using molecular self-assembly the final (desired) structure is programmed in the shape and functional groups of the molecules. Self-assembly is referred to as a 'bottom-up' manufacturing technique in contrast to a 'top-down' technique such as lithography where the desired final structure is carved from a larger block of matter. In the speculative vision of molecular nanotechnology, microchips of the future might be made by molecular self-assembly. An advantage to constructing nanostructure using molecular self-assembly for biological materials is that they will degrade back into individual molecules that can be broken down by the body.

DNA nanotechnology

DNA nanotechnology is an area of current research that uses the bottom-up, self-assembly approach for nanotechnological goals. DNA nanotechnology uses the unique molecular recognition properties of DNA and other nucleic acids to create self-assembling branched DNA complexes with useful properties.[9] DNA is thus used as a structural material rather than as a carrier of biological information, to make structures such as two-dimensional periodic lattices (both tile-based as well as using the "DNA origami" method) and three-dimensional structures in the shapes of polyhedra.[10] These DNA structures have also been used to template the assembly of other molecules such as gold nanoparticles[11] and streptavidin proteins.[12]

See also

- Supramolecular assembly
- Supramolecular chemistry

References

[1] F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma & E. W. Meijer (1998). "Self-Complementarity Achieved through Quadruple Hydrogen Bonding". *Angew. Chem. Int. Ed.* **37** (1-2): 75–78. doi: 10.1002/(SICI)1521-3773(19980202)37:1/2<75::AID-ANIE75>3.0.CO;2-R (http://dx.doi.org/10.1002/(SICI)1521-3773(19980202)37:1/2<75::AID-ANIE75>3.0.CO;2-R).

[2] J.-M. Lehn (1988). "Perspectives in Supramolecular Chemistry-From Molecular Recognition towards Molecular Information Processing and Self-Organization". *Angew. Chem. Int. Ed. Engl.* **27** (11): 89–121. doi: 10.1002/anie.198800891 (http://dx.doi.org/10.1002/anie.198800891).

[3] J.-M. Lehn (1990). "Supramolecular Chemistry-Scope and Perspectives: Molecules, Supermolecules, and Molecular Devices (Nobel Lecture)". *Angew. Chem. Int. Ed. Engl.* **29** (11): 1304–1319. doi: 10.1002/anie.199013041 (http://dx.doi.org/10.1002/anie.199013041).

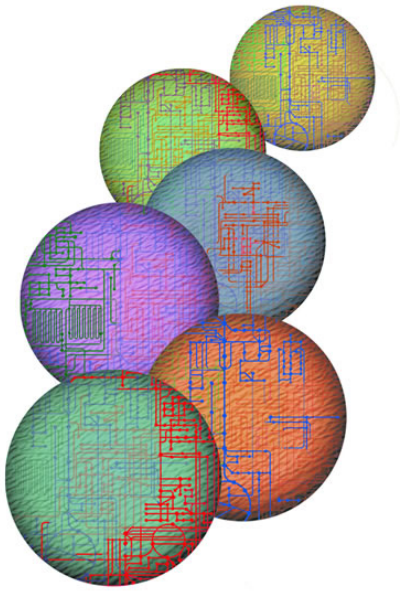
[4] Lehn, J.-M.. *Supramolecular Chemistry: Concepts and Perspectives*. Wiley-VCH. ISBN 978-3-527-29311-7.
[5] Rosen, Milton J. (2004). *Surfactants and interfacial phenomena*. Hoboken, NJ: Wiley-Interscience. ISBN 978-0-471-47818-8.
[6] C. Mao, W. Sun & N. C. Seeman (1997), "Assembly of Borromean rings from DNA", *Nature* **386** (6621): 137–138, doi: 10.1038/386137b0 (http://dx.doi.org/10.1038/386137b0)
[7] K. S. Chichak, S. J. Cantrill, A. R. Pease, S.-H. Chen, G. W. V. Cave, J. L. Atwood & J. F. Stoddart (2004), "Molecular Borromean Rings", *Science* **304** (5675): 1308–1312, doi: 10.1126/science.1096914 (http://dx.doi.org/10.1126/science.1096914), PMID 15166376
[8] M. Strong (2004). "Protein Nanomachines". *PLoS Biol*. **2** (3): e73-e74. doi: 10.1371/journal.pbio.0020073 (http://dx.doi.org/10.1371/journal.pbio.0020073).
[9] N. C. Seeman (2003). "DNA in a material world". *Nature* **421** (6921): 427–431. doi: 10.1038/nature01406 (http://dx.doi.org/10.1038/nature01406).
[10] J. Chen & N. C. Seeman (1991), "http://www.palgrave-journals.com/doifinder/10.1038/350631a0|Synthesis from DNA of a molecule with the connectivity of a cube" (w), *Nature* **350** (6319): 631–633, doi: 10.1038/350631a0 (http://dx.doi.org/10.1038/350631a0), http://www.palgrave-journals.com/doifinder/10.1038/350631a0
[11] C. A. Mirkin, R. L. Letsinger, R. C. Mucic & J. J. Storhoff (1996). "A DNA-based method for rationally assembling nanoparticles into macroscopic materials". *Nature* **382** (6592): 607–609. doi: 10.1038/382607a0 (http://dx.doi.org/10.1038/382607a0).
[12] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif & T. H. Labean (2003), "http://www.sciencemag.org/cgi/content/abstract/301/5641/1882|DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires", *Science* **301** (5641): 1882–1884, doi: 10.1126/science.1089389 (http://dx.doi.org/10.1126/science.1089389), PMID 14512621, http://www.sciencemag.org/cgi/content/abstract/301/5641/1882

External and further reading

- H.E. Hoster, M. Roos, A. Breitruck, C. Meier, K. Tonigold, T. Waldmann, U. Ziener, K. Landfester, R.J. Behm, *Structure Formation in Bis(terpyridine)Derivative Adlayers – Molecule-Substrate vs. Molecule-Molecule Interactions*, Langmuir 23 (2007) 11570
- Molecular Self-Assembly papers (http://www.esi-topics.com/msa/)
- Beyond molecules: Self-assembly of mesoscopic and macroscopic components (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=122665)
- Whitesides, G. M. & Grzyboski, B. (2002) Science 295, 2418-2421.
- Rothemund PWK, Papadakis N, Winfree E (2004) *Algorithmic Self-Assembly of DNA Sierpinski Triangles* (http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0020424). PLoS Biol 2(12)
- C2 Wiki: Self Assembly from a computer programming perspective (http://c2.com/cgi/wiki?SelfAssembly).
- Mohammadzadegan R, Sheikhi MH (2007) *DNA Nano-Gears* (http://www.informaworld.com/openurl?genre=article&issn=0892-7022&volume=33&issue=13&spage=1071) *Molecular Simulation 33(13); 1071-1081.*
- Structure and Dynamics of Organic Nanostructures (http://www.uni-ulm.de/~hhoster/personal/self_assembly.htm)
- Metal organic coordination networks of oligopyridines and Cu on graphite (http://www.uni-ulm.de/~hhoster/personal/metal_organic.htm)
- "Challenges and breakthroughs in recent research on self-assembly" Sci. Technol. Adv. Mater. 9 No 1(2008) 014109 (96 pages) **free download** (http://dx.doi.org/10.1088/1468-6996/9/1/014109)

Cell signaling

Cell signaling is part of a complex system of communication that governs basic cellular activities and coordinates cell actions.[1] The ability of cells to perceive and correctly respond to their microenvironment is the basis of development, tissue repair, and immunity as well as normal tissue homeostasis. Errors in cellular information processing are responsible for diseases such as cancer, autoimmunity, and diabetes. By understanding cell signaling, diseases may be treated effectively and, theoretically, artificial tissues may be yielded.



A complex community of cells.

Traditional work in biology has focused on studying individual parts of cell signaling pathways. Systems biology research helps us to understand the underlying structure of cell signaling networks and how changes in these networks may affect the transmission and flow of information. Such networks are complex systems in their organization and may exhibit a number of emergent properties including bistability and ultrasensitivity. Analysis of cell signaling networks requires a combination of experimental and theoretical approaches including the development and analysis of simulations and modelling.

Unicellular and multicellular organism cell signaling

Cell signaling has been most extensively studied in the context of human diseases and signaling between cells of a single organism. However, cell signaling may also occur between the cells of two different organisms. In many mammals, early embryo cells exchange signals with cells of the uterus.[3] In the human gastrointestinal tract, bacteria exchange signals with each other and with human epithelial and immune system cells.[4] For the yeast *Saccharomyces cerevisiae* during mating, some cells send a peptide signal (mating factor *pheromones*) into their environment. The mating factor peptide may bind to a cell surface receptor on other yeast cells and induce them to prepare for mating.[5]

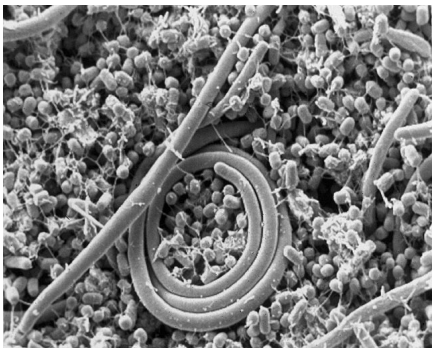


Figure 1. Example of signaling between bacteria. Salmonella enteritidis uses acyl-homoserine lactone for Quorum sensing (see: Inter-Bacterial Communication [2])

Types of signals

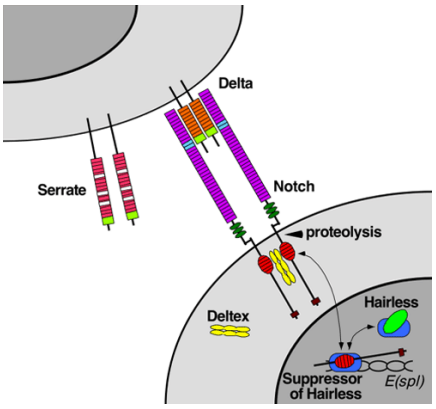


Figure 2. Notch-mediated juxtacrine signal between adjacent cells.

Cells communicate with each other via direct contact (juxtacrine signaling), over short distances (paracrine signaling), or over large distances and/or scales (endocrine signaling).

Some cell-to-cell communication requires direct cell-cell contact. Some cells can form gap junctions that connect their cytoplasm to the cytoplasm of adjacent cells. In cardiac muscle, gap junctions between adjacent cells allows for action potential propagation from the cardiac pacemaker region of the heart to spread and coordinately cause contraction of the heart.

The Notch signaling mechanism is an example of juxtacrine signalling (also known as contact dependent signaling) in which two adjacent cells must make physical contact in order to communicate. This requirement for direct contact allows for very precise control of cell differentiation during embryonic development. In the worm *Caenorhabditis elegans*, two cells of the developing gonad each have an equal chance of terminally differentiating or becoming a uterine precursor cell that continues to divide. The choice of which cell continues to divide is controlled by competition of cell surface signals. One cell will happen to produce more of a cell surface protein that activates the Notch receptor on the adjacent cell. This activates a feedback loop or system that reduces Notch expression in the cell that will differentiate and increases Notch on the surface of the cell that continues as a stem cell.[6]

Many cell signals are carried by molecules that are released by one cell and move to make contact with another cell. *Endocrine* signals are called hormones. Hormones are produced by endocrine cells and they travel through the blood to reach all parts of the body. Specificity of signaling can be controlled if only some cells can respond to a particular hormone. *Paracrine* signals target only cells in the vicinity of the emitting cell. Neurotransmitters represent an example. Some signaling molecules can function as both a hormone and a neurotransmitter. For example, epinephrine and norepinephrine can function as hormones when released from the adrenal gland and are transported to the heart by way of the blood stream. Norepinephrine can also be produced by neurons to function as a neurotransmitter within the brain.[7] Estrogen can be released by the ovary and function as a hormone or act locally via paracrine or autocrine signaling.[8]

Receptors for cell signals

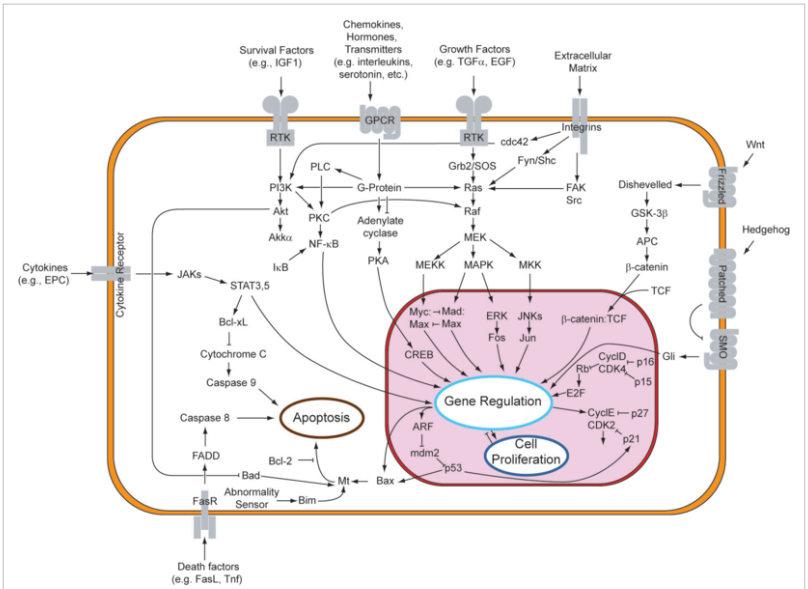
Cells receive information from their environment through a class of proteins known as receptors. Notch is a cell surface protein that functions as a receptor. Animals have a small set of genes that code for signaling proteins that interact specifically with Notch receptors and stimulate a response in cells that express Notch on their surface. Molecules that activate (or, in some cases, inhibit) receptors can be classified as hormones, neurotransmitters, cytokines, growth factors but all of these are called receptor ligands. The details of ligand-receptor interactions are fundamental to cell signaling.

As shown in Figure 2 (above, left), Notch acts as a receptor for ligands that are expressed on adjacent cells. While many receptors are cell surface proteins, some are found inside

cells. For example, estrogen is a hydrophobic molecule that can pass through the lipid bilayer of cell surface membranes. Estrogen receptors inside cells of the uterus can be activated by estrogen that comes from the ovaries, enters the target cells, and binds to estrogen receptors.

Other signaling molecules are unable to permeate the hydrophobic cell membrane due to their hydrophilic nature, so their target receptor is expressed on the membrane. When such signaling molecule activates its receptor, the signal is carried into the cell usually by means of a second messenger such as cAMP.

Signaling pathways



Overview of signal transduction pathways.

In some cases, receptor activation caused by ligand binding to a receptor is directly coupled to the cell's response to the ligand. For example, the neurotransmitter GABA can activate a cell surface receptor that is part of an ion channel. GABA binding to a GABA A receptor on a neuron opens a chloride-selective ion channel that is part of the receptor. GABA A receptor activation allows negatively-charged chloride ions to move into the neuron, which inhibits the ability of the neuron to produce action potentials. However, for many cell surface receptors, ligand-receptor interactions are not directly linked to the cell's response. The activated receptor must first interact with other proteins inside the cell before the ultimate physiological effect of the ligand on the cell's behavior is produced. Often, the behavior of a chain of several interacting cell proteins is altered following receptor activation. The entire set of cell changes induced by

receptor activation is called a signal transduction mechanism or pathway.

In the case of Notch-mediated signaling, the signal transduction mechanism can be relatively simple. As shown in Figure 2 (above, left), activation of Notch can cause the Notch protein to be altered by a protease. Part of the Notch protein is released from the cell surface membrane and can act to change the pattern of gene transcription in the cell nucleus. This causes the responding cell to make different proteins, resulting in an altered pattern of cell behavior. Cell signaling research involves studying the spatial and temporal dynamics of both receptors and the components of signaling pathways that are activated by receptors in various cell types.

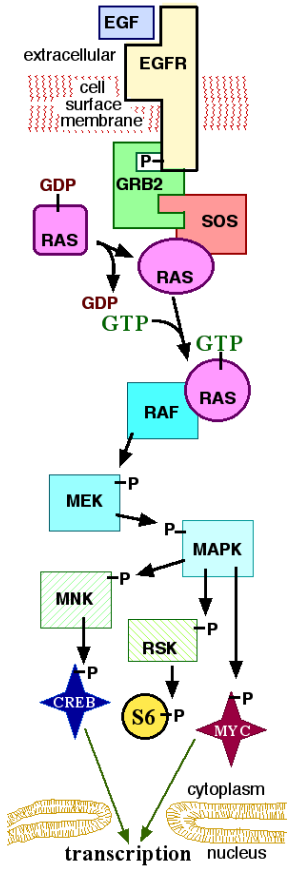


Figure 3. Diagram showing key components of a signal transduction pathway. See the MAPK/ERK pathway article for details.

A more complex signal transduction pathway is shown in Figure 3. This pathway involves changes of protein-protein interactions inside the cell, induced by an external signal. Many growth factors bind to receptors at the cell surface and stimulate cells to progress through the cell cycle and divide. Several of these receptors are kinases that start to phosphorylate themselves and other proteins when binding to a ligand. This phosphorylation can generate a binding site for a different protein and thus induce protein-protein interaction. In Figure 3, the ligand (called epidermal growth factor (EGF)) binds to the receptor (called EGFR). This activates the receptor to phosphorylate itself. The phosphorylated receptor binds to an adaptor protein (GRB2), which couples the signal to further downstream signaling processes. For example, one of the signal transduction pathways that are activated is called the mitogen-activated protein kinase (MAPK) pathway. The signal transduction component labeled as "MAPK" in the pathway was originally called "ERK," so the pathway is called the MAPK/ERK pathway. The MAPK protein is an enzyme, a protein kinase that can attach phosphate to target proteins such as the transcription factor MYC and, thus, alter gene transcription and, ultimately, cell cycle progression. Many cellular proteins are activated downstream of the growth factor receptors (such as EGFR) that initiate this signal transduction pathway.

Some signaling transduction pathways respond differently depending on the amount of signaling received by the cell. For instance, the hedgehog protein activates different genes, depending on the amount of hedgehog protein present.

Complex multi-component signal transduction pathways provide opportunities for feedback, signal amplification, and interactions inside one cell between multiple signals and signaling pathways.

Classification of intercellular communication

Within endocrinology (the study of intercellular signalling in animals) and the endocrine system, intercellular signalling is subdivided into the following classifications:

- *Endocrine* signals are produced by endocrine cells and travel through the blood to reach all parts of the body.
- *Paracrine* signals target only cells in the vicinity of the emitting cell. Neurotransmitters represent an example.
- *Autocrine* signals affect only cells that are of the same cell type as the emitting cell. An example for autocrine signals is found in immune cells.
- *Juxtacrine* signals are transmitted along cell membranes via protein or lipid components integral to the membrane and are capable of affecting either the emitting cell or cells immediately adjacent.

See also

- Molecular Cellular Cognition
- Crosstalk (biology)
- MAPK signaling pathway
- Hedgehog signaling pathway
- TGF beta signaling pathway
- JAK-STAT signaling pathway
- cAMP dependent pathway
- Signal transduction
- Systems biology
- Semiotics
- Lipid signaling

References

[1] Witzany, G.111 (2000). Life: The Communicative Structure. Norderstedt, Libri BoD.

[2] http://www.ars.usda.gov/is/AR/archive/jan00/bact0100.htm

[3] O. A. Mohamed, M. Jonnaert, C. Labelle-Dumais, K. Kuroda, H. J. Clarke and D. Dufort (2005) "Uterine Wnt/beta-catenin signaling is required for implantation" in *Proceedings of the National Academy of Sciences of the United States of America* Volume 102, pages 8579-8584. Entrez Pubmed 15930138 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15930138).

[4] M.B. Clarke and V. Sperandio (2005) "Events at the host-microbial interface of the gastrointestinal tract III. Cell-to-cell signaling among microbial flora, poop, and pathogens: there is a whole lot of talking going on" in *American journal of physiology. Gastrointestinal and liver physiology.* Volume 288, pages G1105-9. Entrez Pubmed 15890712 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15890712).

[5] J. C. Lin, K. Duell and J. B. Konopka (2004) "A microdomain formed by the extracellular ends of the transmembrane domains promotes activation of the G protein-coupled alpha-factor receptor" in *Molecular Cell Biology* Volume 24, pages 2041-2051. Entrez Pubmed 14966283 (http://www.ncbi.nlm.nih.gov/entrez/

query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=14966283).
[6] I. Greenwald (1998) "LIN-12/Notch signaling: lessons from worms and flies" in *Genes in Development* Volume 12, pages 1751-1762. Entrez Pubmed 9637676 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=9637676).
[7] M. C. Cartford, A. Samec, M. Fister and P. C. Bickford (2004) "Cerebellar norepinephrine modulates learning of delay classical eyeblink conditioning: evidence for post-synaptic signaling via PKA" in *Learning & memory* Volume 11, pages 732-737. Entrez Pubmed 15537737 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15537737).
[8] S. Jesmin, C. N. Mowa, I. Sakuma, N. Matsuda, H. Togashi, M. Yoshioka, Y. Hattori and A. Kitabatake (2004) "Aromatase is abundantly expressed by neonatal rat penis but downregulated in adulthood" in *Journal of Molecular Endocrinology* Volume 33, pages 343-359. Entrez Pubmed 15525594 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15525594).

External links

- Signaling Gateway (http://www.signaling-gateway.org) Free summaries of recent research and the Molecule Pages database (http://www.signaling-gateway.org/molecule/).
- NCI-Nature Pathway Interaction Database (http://pid.nci.nih.gov): authoritative information about signaling pathways in human cells.
- Cell Communication (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term="Cell+signaling"+AND+mboc4[book]+AND+373842[uid]&rid=mboc4.section.2743#2793), Chapter 15 in *Molecular Biology of the Cell* (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=cell+biology+AND+mboc4[book]+AND+373693[uid]&rid=mboc4) fourth edition, edited by Bruce Alberts (2002) published by Garland Science.
- Cell Signaling (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term="Cell+signaling"+AND+cooper[book]+AND+166039[uid]&rid=cooper.chapter.2198), Chapter 13 in *The Cell - A Molecular Approach* (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=cell+biology+AND+cooper[book]+AND+165077[uid]&rid=cooper.chapter.89) second edition, by Geoffrey M. Cooper (2000) published by Sinauer Associates.
- Cell-to-Cell Signaling (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term="Cell+signaling"+AND+mcb[book]+AND+107116[uid]&rid=mcb.chapter.5687), Chapter 20 in *Molecular Cell Biology* (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=cell+biology+AND+mcb[book]+AND+105032[uid]&rid=mcb.chapter.145) fourth edition, edited by Harvey Lodish (2000) published by W. H. Freeman and Company.
- MeSH *Intercellular+Signaling+Peptides+and+Proteins* (http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi?mode=&term=Intercellular+Signaling+Peptides+and+Proteins)
- MeSH *Cell+Communication* (http://www.nlm.nih.gov/cgi/mesh/2009/MB_cgi?mode=&term=Cell+Communication)
- ESIGNET Research Project (http://www.esignet.net)
- International q-bio Conference on Cellular Information Processing

Molecular evolution

Molecular evolution is the process of evolution at the scale of DNA, RNA, and proteins. Molecular evolution emerged as a scientific field in the 1960s as researchers from molecular biology, evolutionary biology and population genetics sought to understand recent discoveries on the structure and function of nucleic acids and protein. Some of the key topics that spurred development of the field have been the evolution of enzyme function, the use of nucleic acid divergence as a "molecular clock" to study species divergence, and the origin of non-functional or junk DNA. Recent advances in genomics, including whole-genome sequencing, high-throughput protein characterization, and bioinformatics have led to a dramatic increase in studies on the topic. In the 2000s, some of the active topics have been the role of gene duplication in the emergence of novel gene function, the extent of adaptive molecular evolution versus neutral drift, and the identification of molecular changes responsible for various human characteristics especially those pertaining to infection, disease, and cognition.

Principles of molecular evolution

Mutations

Mutations are permanent, transmissible changes to the genetic material (usually DNA or RNA) of a cell. Mutations can be caused by copying errors in the genetic material during cell division and by exposure to radiation, chemicals, or viruses, or can occur deliberately under cellular control during the processes such as meiosis or hypermutation. Mutations are considered the driving force of evolution, where less favorable (or *deleterious*) mutations are removed from the gene pool by natural selection, while more favorable (or *beneficial*) ones tend to accumulate. Neutral mutations do not affect the organism's chances of survival in its natural environment and can accumulate over time, which might result in what is known as punctuated equilibrium; the modern interpretation of classic evolutionary theory.

Causes of change in allele frequency

There are three known processes that affect the survival of a characteristic; or, more specifically, the frequency of an allele (variant of a gene):

- Genetic drift describes changes in gene frequency that cannot be ascribed to selective pressures, but are due instead to events that are unrelated to inherited traits. This is especially important in small mating populations, which simply cannot have enough offspring to maintain the same gene distribution as the parental generation.
- Gene flow or Migration: or gene admixture is the only one of the agents that makes populations closer genetically while building larger gene pools.
- Selection, in particular natural selection produced by differential mortality and fertility. Differential mortality is the survival rate of individuals before their reproductive age. If they survive, they are then selected further by differential fertility – that is, their total genetic contribution to the next generation. In this way, the alleles that these surviving individuals contribute to the gene pool will increase the frequency of those alleles. Sexual selection, the attraction between mates that results from two genes, one for a feature and the other determining a preference for that feature, is also very important.

Molecular study of phylogeny

Molecular systematics is a product of the traditional field of systematics and molecular genetics. It is the process of using data on the molecular constitution of biological organisms' DNA, RNA, or both, in order to resolve questions in systematics, i.e. about their correct scientific classification or taxonomy from the point of view of evolutionary biology.

Molecular systematics has been made possible by the availability of techniques for DNA sequencing, which allow the determination of the exact sequence of nucleotides or *bases* in either DNA or RNA. At present it is still a long and expensive process to sequence the entire genome of an organism, and this has been done for only a few species. However, it is quite feasible to determine the sequence of a defined area of a particular chromosome. Typical molecular systematic analyses require the sequencing of around 1000 base pairs.

The driving forces of evolution

Depending on the relative importance assigned to the various forces of evolution, three perspectives provide evolutionary explanations for molecular evolution.[1]

While recognizing the importance of random drift for silent mutations,[2] **selectionists hypotheses** argue that balancing and positive selection are the driving forces of molecular evolution. Those hypotheses are often based on the broader view called panselectionism, the idea that selection is the only force strong enough to explain evolution, relaying random drift and mutations to minor roles.[1]

Neutralists hypotheses emphasize the importance of mutation, purifying selection and random genetic drift.[3] The introduction of the neutral theory by Kimura,[4] quickly followed by King and Jukes' own findings,[5] lead to a fierce debate about the relevance of neodarwinism at the molecular level. The Neutral theory of molecular evolution states that most mutations are deleterious and quickly removed by natural selection, but of the remaining ones, the vast majority are neutral with respect to fitness while the amount of advantageous mutations is vanishingly small. The fate of neutral mutations are governed by genetic drift, and contribute to both nucleotide polymorphism and fixed differences between species. [6] [7] [8]

Mutationists hypotheses emphasize random drift and biases in mutation patterns.[9] Sueoka was the first to propose a modern mutationist view. He proposed that the variation in GC content was not the result of positive selection, but a consequence of the GC mutational pressure.[10]

Related fields

An important area within the study of molecular evolution is the use of molecular data to determine the correct biological classification of organisms. This is called molecular systematics or molecular phylogenetics.

Tools and concepts developed in the study of molecular evolution are now commonly used for comparative genomics and molecular genetics, while the influx of new data from these fields has been spurring advancement in molecular evolution.

Key researchers in molecular evolution

Some researchers who have made key contributions to the development of the field:

- Motoo Kimura — Neutral theory
- Masatoshi Nei — Adaptive evolution
- Walter M. Fitch — Phylogenetic reconstruction
- Walter Gilbert — RNA world
- Joe Felsenstein — Phylogenetic methods
- Susumu Ohno — Gene duplication
- John H. Gillespie — Mathematics of adaptation

Journals and societies

Journals dedicated to molecular evolution include *Molecular Biology and Evolution*, *Journal of Molecular Evolution*, and *Molecular Phylogenetics and Evolution*. Research in molecular evolution is also published in journals of genetics, molecular biology, genomics, systematics, or evolutionary biology. The Society for Molecular Biology and Evolution [11] publishes the journal "Molecular Biology and Evolution" and holds an annual international meeting.

See also

- History of molecular evolution
- Chemical evolution
- Evolution
- Genetic drift
- *E. coli* long-term evolution experiment
- Evolutionary physiology
- Neutral theory of molecular evolution
- Nucleotide diversity
- Parsimony
- Population genetics
- Selection
- Genomic organization
- Horizontal gene transfer
- Human evolution
- Molecular clock
- Comparative phylogenetics

Further reading

- Li, W.-H. (2006). *Molecular Evolution*. Sinauer. ISBN 0878934804.
- Lynch, M. (2007). *The Origins of Genome Architecture*. Sinauer. ISBN 0878934847.

References

[1] Graur, D. and Li, W.-H. (2000). *Fundamentals of molecular evolution*. Sinauer.

[2] Gillespie, J. H (1991). *The Causes of Molecular Evolution*. Oxford University Press, New York. ISBN 0-19-506883-1.

[3] Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge. ISBN 0-521-23109-4.

[4] Kimura, Motoo (1968). "http://www2.hawaii.edu/~khayes/Journal_Club/fall2006/Kimura_1968_Nature.pdf|Evolutionary rate at the molecular level". *Nature* **217**: 624–626. doi: 10.1038/217624a0 (http://dx.doi.org/10.1038/217624a0). http://www2.hawaii.edu/~khayes/Journal_Club/fall2006/Kimura_1968_Nature.pdf.

[5] King, J.L. and Jukes, T.H. (1969). "http://www.blackwellpublishing.com/ridley/classictexts/king.pdf|Non-Darwinian Evolution". *Science* **164**: 788–798. doi: 10.1126/science.164.3881.788 (http://dx.doi.org/10.1126/science.164.3881.788). PMID 5767777. http://www.blackwellpublishing.com/ridley/classictexts/king.pdf.

[6] Nachman M. (2006). "*Detecting selection at the molecular level*" *in: Evolutionary Genetics: concepts and case studies*. pp. 103–118.

[7] The nearly neutral theory expanded the neutralist perspective, suggesting that several mutations are nearly neutral, which means both random drift and natural selection is relevant to their dynamics.

[8] Ohta, T (1992). "The nearly neutral theory of molecular evolution". *Annual Review of Ecology and Systematics* **23**: 263–286. doi: 10.1146/annurev.es.23.110192.001403 (http://dx.doi.org/10.1146/annurev.es.23.110192.001403).

[9] Nei, M. (2005). "Selectionism and Neutralism in Molecular Evolution". *Molecular Biology and Evolution* **22(12)**: 2318–2342. doi: 10.1093/molbev/msi242 (http://dx.doi.org/10.1093/molbev/msi242). PMID 16120807.

[10] Sueoka, N. (1964). "On the evolution of informational macromolecules". in In: Bryson, V. and Vogel, H.J.. *Evolving genes and proteins*. Academic Press, New-York. pp. 479–496.

[11] http://www.smbe.org

Molecular phylogenetics

Molecular phylogenetics, also known as **molecular systematics**, is the use of the structure of molecules to gain information on an organism's evolutionary relationships. The result of a molecular phylogenetic analysis is expressed in a phylogenetic tree.

Techniques and applications

Every living organism contains DNA, RNA, and proteins. Closely related organisms generally have a high degree of agreement in the molecular structure of these substances, while the molecules of organisms distantly related usually show a pattern of dissimilarity. Conserved sequences such mitochondrial DNA are expected to accumulate mutations over time, and assuming a constant rate of mutation provide a molecular clock for dating divergence. Molecular phylogeny uses such data to build a "relationship tree" that shows the probable evolution of various organisms. Not until recent decades, however, has it been possible to isolate and identify these molecular structures.

The most common approach is the comparison of sequences for genes using sequence alignment techniques to identify similarity. Another application of molecular phylogeny is in DNA barcoding, where the species of an individual organism is identified using small sections of mitochondrial DNA. Another application of the techniques that make this possible can be seen in the very limited field of human genetics, such as the ever more popular use of genetic testing to determine a child's paternity, as well as the emergence of a new branch of criminal forensics focused on evidence known as genetic fingerprinting.

The effect on traditional biological classification schemes in the biological sciences has been dramatic as well. Work that was once immensely labor- and materials-intensive can now be done quickly and easily, leading to yet another source of information becoming available for systematic and taxonomic appraisal. This particular kind of data has become so popular that taxonomical schemes based solely on molecular data may be encountered.

Theoretical background

Early attempts at molecular systematics were also termed as chemotaxonomy and made use of proteins, enzymes, carbohydrates and other molecules which were separated and characterized using techniques such as chromatography. These have been largely replaced in recent times by DNA sequencing which produces the exact sequences of nucleotides or *bases* in either DNA or RNA segments extracted using different techniques. These are generally considered superior for evolutionary studies since the actions of evolution are ultimately reflected in the genetic sequences. At present it is still a long and expensive process to sequence the entire DNA of an organism (its genome), and this has been done for only a few species. However it is quite feasible to determine the sequence of a defined area of a particular chromosome. Typical molecular systematic analyses require the sequencing of around 1000 base pairs. At any location within such a sequence, the bases found in a given position may vary between organisms. The particular sequence found in a given organism is referred to as its haplotype. In principle, since there are four base types, with 1000 base pairs, we could have 4^{1000} distinct haplotypes. However, for organisms within a particular species or in a group of related species, it has been found empirically that only a minority of sites show any variation at all and most of the variations that are

found are correlated, so that the number of distinct haplotypes that are found is relatively small.

In a molecular systematic analysis, the haplotypes are determined for a defined area of genetic material; ideally a substantial sample of individuals of the target species or other taxon are used however many current studies are based on single individuals. Haplotypes of individuals of closely related, but supposedly different, taxa are also determined. Finally, haplotypes from a smaller number of individuals from a definitely different taxon are determined: these are referred to as an *out group*. The base sequences for the haplotypes are then compared. In the simplest case, the difference between two haplotypes is assessed by counting the number of locations where they have different bases: this is referred to as the number of *substitutions* (other kinds of differences between haplotypes can also occur, for example the *insertion* of a section of nucleic acid in one haplotype that is not present in another). Usually the difference between organisms is re-expressed as a *percentage divergence*, by dividing the number of substitutions by the number of base pairs analysed: the hope is that this measure will be independent of the location and length of the section of DNA that is sequenced.

An older and superseded approach was to determine the divergences between the genotypes of individuals by DNA-DNA hybridisation. The advantage claimed for using hybridisation rather than gene sequencing was that it was based on the entire genotype, rather than on particular sections of DNA. Modern sequence comparison techniques overcome this objection by the use of multiple sequences.

Once the divergences between all pairs of samples have been determined, the resulting triangular matrix of differences is submitted to some form of statistical cluster analysis, and the resulting dendrogram is examined in order to see whether the samples cluster in the way that would be expected from current ideas about the taxonomy of the group, or not. Any group of haplotypes that are all more similar to one another than any of them is to any other haplotype may be said to constitute a clade. Statistical techniques such as bootstrapping and jackknifing help in providing reliability estimates for the positions of haplotypes within the evolutionary trees.

Characteristics and assumptions of molecular systematics

This example illustrates several characteristics of molecular systematics and its underlying assumptions.

1. Molecular systematics is an essentially cladistic approach: it assumes that classification must correspond to phylogenetic descent, and that all valid taxa must be monophyletic.
2. Molecular systematics often uses the molecular clock assumption that quantitative similarity of genotype is a sufficient measure of the recency of genetic divergence. Particularly in relation to speciation, this assumption could be wrong if either
    1. some relatively small genotypic modification acted to prevent interbreeding between two groups of organisms, or
    2. in different subgroups of the organisms being considered, genetic modification proceeded at different rates.
3. In animals, it is often convenient to use mitochondrial DNA for molecular systematic analysis. However, because in mammals mitochondria are inherited only from the mother, this is not fully satisfactory, because inheritance in the paternal line might not be detected: in the example above, Vilà et al. cite more limited studies with chromosomal

DNA that support their conclusions.

These characteristics and assumptions are not wholly uncontroversial among biological systematists. As a cladistic method, molecular systematics is open to the same criticisms as cladistics in general. It can also be argued that it is a mistake to replace a classification based on visible and ecologically relevant characteristics by one based on genetic details that may not even be expressed in the phenotype. However the molecular approach to systematics, and its underlying assumptions, are gaining increasing acceptance. As gene sequencing becomes easier and cheaper, molecular systematics is being applied to more and more groups, and in some cases is leading to radical revisions of accepted taxonomies.

History of molecular phylogenetics

The theoretical frameworks for molecular systematics were laid in the 1960s in the works of Emile Zuckerkandl, Emanuel Margoliash, Linus Pauling and Walter M. Fitch.[1] Applications of molecular systematics were pioneered by Charles G. Sibley (birds), Herbert C. Dessauer (herpetology), and Morris Goodman (primates), followed by Allan C. Wilson, Robert K. Selander, and John C. Avise (who studied various groups). Work with protein electrophoresis began around 1956. Although the results were not quantitative and did not initially improve on morphological classification, they provided tantalizing hints that long-held notions of the classifications of birds, for example, needed substantial revision. In the period of 1974–1986, DNA-DNA hybridization was the dominant technique.[2]

Further reading

- Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer Associates Incorporated. ISBN 0-87893-177-5.
- Hillis, D. M. & Moritz, C. 1996. *Molecular systematics*. 2nd ed. Sinauer Associates Incorporated. ISBN 0-87893-282-8.
- Page, R. D. M. & Holmes, E. C. 1998. *Molecular evolution: a phylogenetic approach*. Blackwell Science, Oxford. ISBN 0-86542-889-1.

References

[1] Edna Suárez-Díaz & Victor H. Anaya-Muñoz (2008) History, objectivity, and the construction of molecular phylogenies. Stud. Hist. Phil. Biol. & Biomed. Sci. 39:451–468

[2] Ahlquist, Jon E., 1999: Charles G. Sibley: A commentary on 30 years of collaboration. *The Auk*, vol. 116, no. 3 (July 1999). A PDF or DjVu version of this article can be downloaded from the issue's table of contents page (http://elibrary.unm.edu/sora/Auk/v116n03/index.php).

See also

- molecular evolution
- computational phylogenetics
- PhyloCode

External links

- NCBI - Systematics and Molecular Phylogenetics (http://www.ncbi.nlm.nih.gov/About/primer/phylo.html)

Computational phylogenetics

Computational phylogenetics is the application of computational algorithms, methods and programs to phylogenetic analyses. The goal is to assemble a phylogenetic tree representing a hypothesis about the evolutionary ancestry of a set of genes, species, or other taxa. For example, these techniques have been used to explore the family tree of hominid species[1] and the relationships between specific genes shared by many types of organisms.[2] Traditional phylogenetics relies on morphological data obtained by measuring and quantifying the phenotypic properties of representative organisms, while the more recent field of molecular phylogenetics uses nucleotide sequences encoding genes or amino acid sequences encoding proteins as the basis for classification. Many forms of molecular phylogenetics are closely related to and make extensive use of sequence alignment in constructing and refining phylogenetic trees, which are used to classify the evolutionary relationships between homologous genes represented in the genomes of divergent species. The phylogenetic trees constructed by computational methods are unlikely to perfectly reproduce the evolutionary tree that represents the historical relationships between the species being analyzed. The historical species tree may also differ from the historical tree of an individual homologous gene shared by those species.

Producing a phylogenetic tree requires a measure of homology among the characteristics shared by the taxa being compared. In morphological studies, this requires explicit decisions about which physical characteristics to measure and how to use them to encode distinct states corresponding to the input taxa. In molecular studies, a primary problem is in producing a multiple sequence alignment (MSA) between the genes or amino acid sequences of interest. Progressive sequence alignment methods produce a phylogenetic tree by necessity because they incorporate new sequences into the calculated alignment in order of genetic distance. Although a phylogenetic tree can always be constructed from an MSA, phylogenetics methods such as maximum parsimony and maximum likelihood do not require the production of an initial or concurrent MSA.

Types of phylogenetic trees

Phylogenetic trees generated by computational phylogenetics can be either *rooted* or *unrooted* depending on the input data and the algorithm used. A rooted tree is a directed graph that explicitly identifies a most recent common ancestor (MRCA), usually an imputed sequence that is not represented in the input. Genetic distance measures can be used to plot a tree with the input sequences as leaf nodes and their distances from the root proportional to their genetic distance from the hypothesized MRCA. Identification of a root usually requires the inclusion in the input data of at least one "outgroup" known to be only distantly related to the sequences of interest.

By contrast, unrooted trees plot the distances and relationships between input sequences without making assumptions regarding their descent. An unrooted tree can always be produced from a rooted tree, but a root cannot usually be placed on an unrooted tree

without additional data on divergence rates, such as the assumption of the molecular clock hypothesis.[3]

The set of all possible phylogenetic trees for a given group of input sequences can be conceptualized as a discretely defined multidimensional "tree space" through which search paths can be traced by optimization algorithms. Although counting the total number of trees for a nontrivial number of input sequences can be complicated by variations in the definition of a tree topology, it is always true that there are more rooted than unrooted trees for a given number of inputs and choice of parameters.[4]

Coding characters and defining homology

Morphological analysis

The basic problem in morphological phylogenetics is the assembly of a matrix representing a mapping from each of the taxa being compared to representative measurements for each of the phenotypic characteristics being used as a classifier. The types of phenotypic data used to construct this matrix depend on the taxa being compared; for individual species, they may involve measurements of average body size, lengths or sizes of particular bones or other physical features, or even behavioral manifestations. Of course, since not every possible phenotypic characteristic could be measured and encoded for analysis, the selection of which features to measure is a major inherent obstacle to the method. The decision of which traits to use as a basis for the matrix necessarily represents a hypothesis about which traits of a species or higher taxon are evolutionarily relevant.[5] Morphological studies can be confounded by examples of convergent evolution of phenotypes.[6] A major challenge in constructing useful classes is the high likelihood of inter-taxon overlap in the distribution of the phenotype's variation. The inclusion of extinct taxa in morphological analysis is often difficult due to absence of or incomplete fossil records, but has been shown to have a significant effect on the trees produced; in one study only the inclusion of extinct species of apes produced a morphologically derived tree that was consistent with that produced from molecular data.[1]

Some phenotypic classifications, particularly those used when analyzing very diverse groups of taxa, are discrete and unambiguous; classifying organisms as possessing or lacking a tail, for example, is straightforward in the majority of cases, as is counting features such as eyes or vertebrae. However, the most appropriate representation of continuously varying phenotypic measurements is a controversial problem without a general solution. A common method is simply to sort the measurements of interest into two or more classes, rendering continuous observed variation as discretely classifiable (e.g., all examples with humerus bones longer than a given cutoff are scored as members of one state, and all members whose humerus bones are shorter than the cutoff are scored as members of a second state). This results in an easily manipulated data set but has been criticized for poor reporting of the basis for the class definitions and for sacrificing information compared to methods that use a continuous weighted distribution of measurements.[7]

Because morphological data is extremely labor-intensive to collect, whether from literature sources or from field observations, reuse of previously compiled data matrices is not uncommon, although this may propagate flaws in the original matrix into multiple derivative analyses.[8]

Molecular analysis

The problem of character coding is very different in molecular analyses, as the characters in biological sequence data are immediate and discretely defined - distinct nucleotides in DNA or RNA sequences and distinct amino acids in protein sequences. However, defining homology can be challenging due to the inherent difficulties of multiple sequence alignment. For a given gapped MSA, several rooted phylogenetic trees can be constructed that vary in their interpretations of which changes are "mutations" versus ancestral characters, and which events are insertion mutations or deletion mutations. For example, given only a pairwise alignment with a gap region, it is impossible to determine whether one sequence bears an insertion mutation or the other carries a deletion. The problem is magnified in MSAs with unaligned and nonoverlapping gaps. In practice, sizable regions of a calculated alignment may be discounted in phylogenetic tree construction to avoid integrating noisy data into the tree calculation.

Distance-matrix methods

Distance-matrix methods of phylogenetic analysis explicitly rely on a measure of "genetic distance" between the sequences being classified, and therefore they require an MSA as an input. Distance is often defined as the fraction of mismatches at aligned positions, with gaps either ignored or counted as mismatches.[3] Distance methods attempt to construct an all-to-all matrix from the sequence query set describing the distance between each sequence pair. From this is constructed a phylogenetic tree that places closely related sequences under the same interior node and whose branch lengths closely reproduce the observed distances between sequences. Distance-matrix methods may produce either rooted or unrooted trees, depending on the algorithm used to calculate them. They are frequently used as the basis for progressive and iterative types of multiple sequence alignments. The main disadvantage of distance-matrix methods is their inability to efficiently use information about local high-variation regions that appear across multiple subtrees.[4]

Neighbor-joining

Neighbor-joining methods apply general data clustering techniques to sequence analysis using genetic distance as a clustering metric. The simple neighbor-joining method produces unrooted trees, but it does not assume a constant rate of evolution (i.e., a molecular clock) across lineages. Its relative, UPGMA (Unweighted Pair Group Method with Arithmetic mean) produces rooted trees and requires a constant-rate assumption - that is, it assumes an ultrametric tree in which the distances from the root to every branch tip are equal.

Fitch-Margoliash method

The Fitch-Margoliash method uses a weighted least squares method for clustering based on genetic distance.[9] Closely related sequences are given more weight in the tree construction process to correct for the increased inaccuracy in measuring distances between distantly related sequences. The distances used as input to the algorithm must be normalized to prevent large artifacts in computing relationships between closely related and distantly related groups. The distances calculated by this method must be linear; the linearity criterion for distances requires that the expected values of the branch lengths for two individual branches must equal the expected value of the sum of the two branch

distances - a property that applies to biological sequences only when they have been corrected for the possibility of back mutations at individual sites. This correction is done through the use of a substitution matrix such as that derived from the Jukes-Cantor model of DNA evolution. The distance correction is only necessary in practice when the evolution rates differ among branches.[4]

The least-squares criterion applied to these distances is more accurate but less efficient than the neighbor-joining methods. An additional improvement that corrects for correlations between distances that arise from many closely related sequences in the data set can also be applied at increased computational cost. Finding the optimal least-squares tree with any correction factor is NP-complete,[10] so heuristic search methods like those used in maximum-parsimony analysis are applied to the search through tree space.

Using outgroups

Independent information about the relationship between sequences or groups can be used to help reduce the tree search space and root unrooted trees. Standard usage of distance-matrix methods involves the inclusion of at least one outgroup sequence known to be only distantly related to the sequences of interest in the query set.[3] This usage can be seen as a type of experimental control. If the outgroup has been appropriately chosen, it will have a much greater genetic distance and thus a longer branch length than any other sequence, and it will appear near the root of a rooted tree. Choosing an appropriate outgroup requires the selection of a sequence that is moderately related to the sequences of interest; too close a relationship defeats the purpose of the outgroup and too distant adds noise to the analysis.[3] Care should also be taken to avoid situations in which the species from which the sequences were taken are distantly related, but the gene encoded by the sequences is highly conserved across lineages. Horizontal gene transfer, especially between otherwise divergent bacteria, can also confound outgroup usage.

Maximum parsimony

Maximum parsimony (MP) is a method of identifying the potential phylogenetic tree that requires the smallest total number of evolutionary events to explain the observed sequence data. Some ways of scoring trees also include a "cost" associated with particular types of evolutionary events and attempt to locate the tree with the smallest total cost. This is a useful approach in cases where not every possible type of event is equally likely - for example, when particular nucleotides or amino acids are known to be more mutable than others.

The most naive way of identifying the most parsimonious tree is simple enumeration - considering each possible tree in succession and searching for the tree with the smallest score. However, this is only possible for a relatively small number of sequences or species because the problem of identifying the most parsimonious tree is known to be NP-hard;[4] consequently a number of heuristic search methods for optimization have been developed to locate a highly parsimonious tree, if not the most optimal in the set. Most such methods involve a steepest descent-style minimization mechanism operating on a tree rearrangement criterion.

Branch and bound

The branch and bound algorithm is a general method used to increase the efficiency of searches for near-optimal solutions of NP-hard problems first applied to phylogenetics in the early 1980s.[11] Branch and bound is particularly well suited to phylogenetic tree construction because it inherently requires dividing a problem into a tree structure as it subdivides the problem space into smaller regions. As its name implies, it requires as input both a branching rule (in the case of phylogenetics, the addition of the next species or sequence to the tree) and a bound (a rule that excludes certain regions of the search space from consideration, thereby assuming that the optimal solution cannot occupy that region). Identifying a good bound is the most challenging aspect of the algorithm's application to phylogenetics. A simple way of defining the bound is a maximum number of assumed evolutionary changes allowed per tree. A set of criteria known as Zharkikh's rules[12] severely limit the search space by defining characteristics shared by all candidate "most parsimonious" trees. The two most basic rules require the elimination of all but one redundant sequence (for cases where multiple observations have produced identical data) and the elimination of character sites at which two or more states do not occur in at least two species. Under ideal conditions these rules and their associated algorithm would completely define a tree.

Sankoff-Morel-Cedergren algorithm

The Sankoff-Morel-Cedergren algorithm was among the first published methods to simultaneously produce an MSA and a phylogenetic tree for nucleotide sequences.[13] The method uses a maximum parsimony calculation in conjunction with a scoring function that penalizes gaps and mismatches, thereby favoring the tree that introduces a minimal number of such events. The imputed sequences at the interior nodes of the tree are scored and summed over all the nodes in each possible tree. The lowest-scoring tree sum provides both an optimal tree and an optimal MSA given the scoring function. Because the method is highly computationally intensive, an approximate method in which initial guesses for the interior alignments are refined one node at a time. Both the full and the approximate version are in practice calculated by dynamic programming.[4]

MALIGN and POY

More recent phylogenetic tree/MSA methods use heuristics to isolate high-scoring, but not necessarily optimal, trees. The MALIGN method uses a maximum-parsimony technique to compute a multiple alignment by maximizing a cladogram score, and its companion POY uses an iterative method that couples the optimization of the phylogenetic tree with improvements in the corresponding MSA.[14] However, the use of these methods in constructing evolutionary hypotheses has been criticized as biased due to the deliberate construction of trees reflecting minimal evolutionary events.[15] Both programs are available from the American Museum of Natural History [16].

Maximum likelihood

The maximum likelihood method uses standard statistical techniques for inferring probability distributions to assign probabilities to particular possible phylogenetic trees. The method requires a substitution model to assess the probability of particular mutations; roughly, a tree that requires more mutations at interior nodes to explain the observed phylogeny will be assessed as having a lower probability. This is broadly similar to the maximum-parsimony method, but maximum likelihood allows additional statistical flexibility by permitting varying rates of evolution across both lineages and sites. In fact, the method requires that evolution at different sites and along different lineages must be statistically independent. Maximum likelihood is thus well suited to the analysis of distantly related sequences, but because it formally requires search of all possible combinations of tree topology and branch length, it is computationally expensive to perform on more than a few sequences.

The "pruning" algorithm, a variant of dynamic programming, is often used to reduce the search space by efficiently calculating the likelihood of subtrees.[4] The method calculates the likelihood for each site in a "linear" manner, starting at a node whose only descendants are leaves (that is, the tips of the tree) and working backwards toward the "bottom" node in nested sets. However, the trees produced by the method are only rooted if the substitution model is irreversible, which is not generally true of biological systems. The search for the maximum-likelihood tree also includes a branch length optimization component that is difficult to improve upon algorithmically; general global optimization tools such as the Newton-Raphson method are often used. Searching tree topologies defined by likelihood has not been shown to be NP-complete,[4] but remains extremely challenging because branch-and-bound search is not yet effective for trees represented in this way.

Bayesian inference

Bayesian inference can be used to produce phylogenetic trees in a manner closely related to the maximum likelihood methods. Bayesian methods assume a prior probability distribution of the possible trees, which may simply be the probability of any one tree among all the possible trees that could be generated from the data, or may be a more sophisticated estimate derived from the assumption that divergence events such as speciation occur as stochastic processes. The choice of prior distribution is a point of contention among users of Bayesian-inference phylogenetics methods.[4]

Implementations of Bayesian methods generally use Markov chain Monte Carlo sampling algorithms, although the choice of move set varies; selections used in Bayesian phylogenetics include circularly permuting leaf nodes of a proposed tree at each step[17] and swapping descendant subtrees of a random internal node between two related trees.[18] The use of Bayesian methods in phylogenetics has been controversial, largely due to incomplete specification of the choice of move set, acceptance criterion, and prior distribution in published work.[4]

Model selection

Molecular phylogenetics methods rely on a defined substitution model that encodes a hypothesis about the relative rates of mutation at various sites along the gene or amino acid sequences being studied. At their simplest, substitution models aim to correct for differences in the rates of transitions and transversions in nucleotide sequences. The use of substitution models is necessitated by the fact that the genetic distance between two sequences increases linearly only for a short time after the two sequences diverge from each other (alternatively, the distance is linear only shortly before coalescence). The longer the amount of time after divergence, the more likely it becomes that two mutations occur at the same nucleotide site. Simple genetic distance calculations will thus undercount the number of mutation events that have occurred in evolutionary history. The extent of this undercount increases with increasing time since divergence, which can lead to the phenomenon of long branch attraction, or the misassignment of two distantly related but convergently evolving sequences as closely related.[19] The maximum parsimony method is particularly susceptible to this problem due to its explicit search for a tree representing a minimum number of distinct evolutionary events.[4]

Types of models

All substitution models assign a set of weights to each possible change of state represented in the sequence. The most common model types are implicitly reversible because they assign the same weight to, for example, a G>C nucleotide mutation as to a C>G mutation. The simplest possible model, the Jukes-Cantor model, assigns an equal probability to every possible change of state for a given nucleotide base. The rate of change between any two distinct nucleotides will be one-third of the overall substitution rate.[4] More advanced models distinguish between transitions and transversions. The most general possible time-reversible model, called the GTR model, has contains six mutation rate parameters. An even more generalized model known as the general 12-parameter model breaks time-reversibility, at the cost of much additional complexity in calculating genetic distances that are consistent among multiple lineages.[4] One possible variation on this theme adjusts the rates so that overall GC content - an important measure of DNA double helix stability - varies over time.[20]

Models may also allow for the variation of rates with positions in the input sequence. The most obvious example of such variation follows from the arrangement of nucleotides in protein-coding genes into three-base codons. If the location of the open reading frame (ORF) is known, rates of mutation can be adjusted for position of a given site within a codon, since it is known that wobble base pairing can allow for higher mutation rates in the third nucleotide of a given codon without affecting the codon's meaning in the genetic code.[19] A less hypothesis-driven example that does not rely on ORF identification simply assigns to each site a rate randomly drawn from a predetermined distribution, often the gamma distribution or log-normal distribution.[4] Finally, a more conservative estimate of rate variations known as the covarion method allows autocorrelated variations in rates, so that the mutation rate of a given site is correlated across sites and lineages.[21]

Choosing the best model

The selection of an appropriate model is critical for the production of good phylogenetic analyses, both because underparameterized or overly restrictive models may produce aberrant behavior when their underlying assumptions are violated, and because overly complex or overparameterized models are computationally expensive and the parameters may be overfit.[19] The most common method of model selection is the likelihood ratio test (LRT), which produces a likelihood estimate that can be interpreted as a measure of "goodness of fit" between the model and the input data.[19] However, care must be taken in using these results, since a more complex model with more parameters will always have a higher likelihood than a simplified version of the same model, which can lead to the naive selection of models that are overly complex.[4] For this reason model selection computer programs will choose the simplest model that is not significantly worse than more complex substitution models. A significant disadvantage of the LRT is the necessity of making a series of pairwise comparisons between models; it has been shown that the order in which the models are compared has a major effect on the one that is eventually selected.[22]

An alternative model selection method is the Akaike information criterion (AIC), formally an estimate of the Kullback-Leibler divergence between the true model and the model being tested. It can be interpreted as a likelihood estimate with a correction factor to penalize overparameterized models.[19] The AIC is calculated on an individual model rather than a pair, so it is independent of the order in which models are assessed. A related alternative, the Bayesian information criterion (BIC), has a similar basic interpretation but penalizes complex models more heavily.[19]

See also

- List of phylogenetics software
- Cladistics
- PHYLIP
- Phylogenetic comparative methods
- Phylogenetic tree
- Phylogenetics
- Systematics
- Joe Felsenstein

External links

- PHYLIP [23], a freely distributed phylogenetic analysis package
- PAUP [24], a similar analysis package available for purchase
- MrBayes [25], a program for the Bayesian estimation of phylogeny (software wiki [26])
- BAli-Phy [27], a program for simultaneous Bayesian estimation of alignment and phylogeny.
- Treefinder [28], a graphical analysis environment for molecular phylogenetics
- Modeltest [29], a program for selecting appropriate substitution models for nucleotide sequences
- CIPRES: Cyberinfrastructure for Phylogenetic Research [30]
- Phylogenetic inferring on the T-REX server [31]
- List of phylogeny programs [32]
- Phylogeny Algorithms Pseudocode [33]

References

[1] Strait DS, Grine FE. (2004). Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. *J Hum Evol* 47(6):399-452.

[2] Hodge T, Cope MJ. (2000). A myosin family tree. *J Cell Sci* 113: 3353-3354.

[3] Mount DM. (2004). *Bioinformatics: Sequence and Genome Analysis* 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

[4] Felsenstein J. (2004). *Inferring Phylogenies* Sinauer Associates: Sunderland, MA.

[5] Swiderski DL, Zelditch ML, Fink WL. (1998). Why morphometrics is not special: coding quantitative data for phylogenetic analysis. 47(3):508-19.

[6] Gaubert P, Wozencraft WC, Cordeiro-Estrela P, Veron G. (2005). Mosaics of convergences and noise in morphological phylogenies: what's in a viverrid-like carnivoran? *Syst Biol* 54(6):865-94.

[7] Wiens JJ. (2001). Character analysis in morphological phylogenetics: problems and solutions. *Syst Biol* 50(5):689-99.

[8] Jenner RA. (2001). Bilaterian phylogeny and uncritical recycling of morphological data sets. *Syst Biol* 50(5): 730-743.

[9] Fitch WM, Margoliash E. (1967). Construction of phylogenetic trees. *Science* 155: 279-84.

[10] Day, WHE. (1986). Computational complexity of inferring phylogenies from dissimilarity matrices. *Bulletin of Mathematical Biology* 49:461-7.

[11] Hendy MD, Penny D. (1982). Branch and bound algorithms to determine minimal evolutionary trees. *Math Biosci* 60: 133-42.

[12] Ratner VA, Zharkikh AA, Kolchanov N, Rodin S, Solovyov S, Antonov AS. (1995). *Molecular Evolution* Biomathematics Series Vol 24. Springer-Verlag: New York, NY.

[13] Sankoff D, Morel C, Cedergren RJ. (1973). Evolution of 5S RNA and the non-randomness of base replacement. *Nature New Biology* 245:232-4.

[14] Wheeler WC, Gladstein DG. (1994). MALIGN: a multiple nucleic acid sequence alignment program. *J Heredity* 85: 417-18.

[15] Simmons MP. (2004). Independence of alignment and tree search. *Mol Phylogenet Evol* 31(3):874-9.

[16] http://research.amnh.org/scicomp/projects.html

[17] Mau B, Newton MA. (1997). Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J Comp Graph Stat* 6:122-31.

[18] Yang Z, Rannala B. (1997). bayesian phylogenetic inference using DNA sequences: a Merkov chain Monte Carlo method. *Mol Biol Evol* 46:409-18.

[19] Sullivan J, Joyce P. (2005). Model selection in phylogenetics. Annual Review of Ecology, Evolution, and Systematics. 36: 445-466.

[20] Galtier N, Guoy M. (1998.) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol. Biol. Evol.* 15:871–79.

[21] Fitch WM, Markowitz E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics* 4:579-593.

[22] Pol D. (2004.) Empirical problems of the hierarchical likelihood ratio test for model selection. *Syst Biol* 53:949–62.

[23] http://evolution.genetics.washington.edu/phylip.html

[24] http://paup.csit.fsu.edu/

[25] http://mrbayes.csit.fsu.edu/index.php

[26] http://mrbayes.csit.fsu.edu/wiki/index.php/Main_Page

[27] http://www.biomath.ucla.edu/msuchard/bali-phy/

[28] http://www.treefinder.de/

[29] http://darwin.uvigo.es/software/modeltest.html

[30] http://www.phylo.org/

[31] http://www.trex.uqam.ca

[32] http://evolution.genetics.washington.edu/phylip/software.html

[33] http://www.rustyspigot.com/Computer_Science/Bioinformatics.html#phylogeny

Article Sources and Contributors

Molecular modeling *Source*: http://en.wikipedia.org/windex.php?oldid=253168029 *Contributors*: -

Molecular modelling *Source*: http://en.wikipedia.org/windex.php?oldid=293943983 *Contributors*: ACBest, Agilemolecule, Akpakp, Amalas, Anonymi, Anotherdendron, Bbullot, Bduke, BenFrantzDale, Bensaccount, Bikadi, Billgunn, Bio-ITWorld, Biophys, BokicaK, Bspahh, Cacycle, Chenmengen, DMacks, DeadEyeArrow, Dicklyon, Drmolecule, Edguy99, ElaineMeng, Gaius Cornelius, Gogo Dodo, Graik, Hu, Hu12, Icairns, Itub, JaGa, Jamitzky, Jason Quinn, Joerg Kurt Wegner, Karol Langner, Kazkaskazkasako, Keilana, Lantonov, Lexor, LiDaobing, Ms2ger, Nicolazonta, Ohlinger, Ohnoitsjamie, P99am, Para, Petermr, ProteusCoop, Puppy8800, Rich Farmbrough, Shura58, Splette, TheParanoidOne, Transisto, Trevyn, Unconcerned, Utcursch, Uvainio, Van helsing, Vriend, Weiguxp, Wikimcmd, Xebvor, YAYsocialism, Zmm, 96 anonymous edits

Quantum chemistry *Source*: http://en.wikipedia.org/windex.php?oldid=292520271 *Contributors*: 144.189.40.xxx, 208.40.185.xxx, 4lex, Acroterion, Alansohn, Ayla, BTDenyer, Bci2, Bduke, Bob, BrianY, Bubbha, CDN99, Capecodeph, ChemGardener, CloudNine, Cmdrjameson, CommonsDelinker, Conversion script, Cool3, Cypa, EdJohnston, Edsanville, EmilyT, Euryalus, Gentgeen, Gershom, Giftlite, Glenn, GregorB, Haljolad, HappyCamper, Holdran, Hugo-cs, Ian Pitchford, Itub, James 007, Jantop, JerrySteal, Kaliumfredrik, Karol Langner, Keenan Pepper, Keilana, Koinut, Krash, La goutte de pluie, Lampuchi, Ligulem, Lijuni, Looxix, M stone, Martin Hedegaard, Meisfunny, Milo, Nickptar, Noisy, Nzzl, Okedem, Perelaar, Ratol, Rifleman 82, SHL-at-Sv, SQL, Sadi Carnot, Salsb, Shalom Yechiel, Shanel, Sidhekin, Smoe, Sunev, Tasudrty, Terhorstj, Timwi, UninvitedCompany, Vb, Vgy7ujm, Vig vimarsh, Voigfdsa, Vsmith, W.F.Galway, Wiki alf, Yurik, Zarniwoot, Zeimusu, Александър, علاع بوبحم, 132 anonymous edits

Molecular orbital theory *Source*: http://en.wikipedia.org/windex.php?oldid=281768686 *Contributors*: Bduke, Bensaccount, CIreland, Crackles, Cspan64, GLaDOS, Grimlock, HappyCamper, Indon, Itub, J.delanoy, Jaganath, Omegakent, Pelirojopajaro, Puppy8800, Robertfreemanfund, S Levchenkov, Sadi Carnot, Sbisolo, Uri2, V8rik, Wackedout, Withlyn, 18 anonymous edits

Linear combination of atomic orbitals molecular orbital method *Source*: http://en.wikipedia.org/windex.php?oldid=289673534 *Contributors*: BRG, Badanedwa, Bduke, Brews ohare, ChemGrrl, Chmrjg, Cholerashot, Christianjb, Crystal whacker, DMacks, Eequor, Fred Bradstadt, HappyCamper, Hellisp, Karl-Henner, Linas, LuchoX, Marquez, Michael Hardy, Physchim62, Salsb, Topbanana, V8rik, Vb, 28 anonymous edits

Hückel method *Source*: http://en.wikipedia.org/windex.php?oldid=290722301 *Contributors*: Bduke, Brichcja, Bwibbwz, Charles Matthews, Davodd, Essin, FelixP, Fuhghettaboutit, Gene Nygaard, Grimlock, Ian Glenn, Itub, LuchoX, Martin Hedegaard, Mikhailov Kusserow, Mlaffs, Pearle, Pegship, Petenweller, Rjwilmsi, V8rik, Washburnmav, William915, 13 anonymous edits

Extended Huckel method *Source*: http://en.wikipedia.org/windex.php?oldid=268570409 *Contributors*: Andyras, Bduke, Charles Matthews, Chemistrymouse, Everyking, Grimlock, Itub, Jschrier, Qitaana, RFenno, Sj, Tomo, 11 anonymous edits

Molecular graphics *Source*: http://en.wikipedia.org/windex.php?oldid=285442351 *Contributors*: ALoopingIcon, Agilemolecule, Altenmann, Arch dude, Chemistrannik, Chenmengen, CzarB, Dcrjsr, Dreftymac, Edguy99, Edward, EranHodis, Fvasconcellos, Harryboyles, Icep, JLSussman, Jweiss11, Karol Langner, Linforest, McVities, Mdd, Mobius, Mrug2, NapoliRoma, NicoV, Ohnoitsjamie, Outriggr, P99am, PBarak, Petermr, ProveIt, Rjwilmsi, Rogerb67, SchuminWeb, Shura58, Sjoerd de Vries, SkyWalker, Thumperward, Timrollpickering, Vriend, Walkerma, 16 anonymous edits

List of software for molecular mechanics modeling *Source*: http://en.wikipedia.org/windex.php?oldid=293821193 *Contributors*: Agilemolecule, Alma-Tadema, Anxdo, Bduke, Drbreznjev, ElaineMeng, Enric Naval, Gsmith8, Kiplingw, Leevanjackson, Ludx, Mijuva, Mndoci, Mollwollfumble, Nicolazonta, P99am, Pemmy, Radiometer, Sekmi, Sellersb, Shadowboy813, Thorwald, UnitedStatesian, Vi2, Vriend, WikHead, Wrpscott, Xebvor, 51 anonymous edits

Protein structure prediction *Source*: http://en.wikipedia.org/windex.php?oldid=287834213 *Contributors*: 168..., Agilemolecule, Aiminy, Alex.g, Antelan, Aroopsircar, Asasia, Biophys, Blastwizard, Bsiraptor, Chad.davis, Ched Davis, Christo07, CommodiCast, Davjon, Dcooper, De728631, DerHexer, Dhatz, Dmb000006, DoctaDontist, Dulcet86, Emw2012, Figureskatingfan, Fvasconcellos, Gaius Cornelius, Gasqw, Gschizas, Herr blaschke, II MusLiM HyBRiD II, Icarus3, Intangir, Itub, JWSchmidt, JaGa, JavierMC, KaHa242, Kaare, Kevyn, Kierano, Kjaergaard, Kku, Lexor, Lionfish0, Malcolm Farmer, MayDoris, MichaK, Minghong, Nsaa, Opabinia regalis, P99am, Password, Piano non troppo, Pol098, Q31245, Randommouse, RexNL, Samohyl Jan, Selain03, Smuskal, Snowmanradio, SnowyDay, Soeding, Stewartadcock, TenOfAllTrades, TestPilot, Thorwald, Tomixdf, Tregoweth, WhiteDragon, Wik, WillowW, Zargulon, 73 anonymous edits

Protein design *Source*: http://en.wikipedia.org/windex.php?oldid=291199298 *Contributors*: Andraaide, Caltechdoc, Edboas, Gilliam, Huttarl, Ihope127, Lfh, P99am, Rich Farmbrough, Rjwilmsi, Van der Hoorn, Vriend, Zephyris, Zzuuzz, 28 anonymous edits

Homology modeling *Source*: http://en.wikipedia.org/windex.php?oldid=293320720 *Contributors*: Aesopos, Biophys, Blastwizard, Boghog2, Crambin, Cs30109, DanielPenfield, Dewatson, Hemmingsen, Opabinia regalis, Outriggr, P99am, Poccil, QuiteUnusual, R:128.40.76.3, Temiz, Vriend, WillowW, 18 anonymous edits

Loop modeling *Source*: http://en.wikipedia.org/windex.php?oldid=240405982 *Contributors*: Opabinia regalis, WillowW, 2 anonymous edits

MODELLER *Source*: http://en.wikipedia.org/windex.php?oldid=194774445 *Contributors*: Aqwis, Karol Langner, Luk, Opabinia regalis, Tea with milk, Tengku syariful, UnitedStatesian, 1 anonymous edits

Molecular models of DNA *Source*: http://en.wikipedia.org/windex.php?oldid=296022643 *Contributors*: Bci2, Chris the speller, CommonsDelinker, Oscarthecat

List of nucleic acid simulation software *Source*: http://en.wikipedia.org/windex.php?oldid=291597253 *Contributors*: Enric Naval, P99am, Queenie, Thorwald, WikHead, 1 anonymous edits

Folding@home *Source*: http://en.wikipedia.org/windex.php?oldid=293110816 *Contributors*: -Majestic-, 7im, AMK1211, Abdullahazzam, Alfio, Ali@gwc.org.uk, Alton, Amire80, AncientToaster, Ank0ku, Anonymous Dissident, Antisora, Arrenlex, BPinard, Badwolf415, Balmung0731, Bash, Bdesham, BebopBob, Beltz, Bender235, Bensin, Bigboehmboy, Biggins, Bobo192, Bongle, BrOnXbOmBr21, Brian Kendig, Bucetass, Burrito, Cardsplayer4life, Ccson, Ceyockey, ChimpanzeeUK, Chirags, Chowbok, ClementSeveillac, Clicketyclack, CloudNine, Codernaut, Colonies Chris, Compotatoj, Considerinfo, CopperKettle, Copysan, Cowb0y, Credema, CyberSkull, DMay, DOSGuy, Dan100, Dancinginblood, Darkstar1st, Daveswagon, David Cat, Davidmec, Defsac, Deglr6328, Demonkey36, DerHexer, DevastatorIIC, Discospinster, Donald Goldberg, Donarreiskoffer, Drakaal, Drektor2oo3, Drkameleon, Dust Filter, Dwaipayanc, El C, Elapsed, EliasAlucard, Eliot1785, Eouw0o83hf, Epbr123, Eskimo, Ewlyahoocom, Eyu100, Falcor84, FayssalF, FearTec, Feureau, FirefoxRocks, Foundby, Fourchannel, Gaius Cornelius, GalliasM, Goodone121, Gpearson2, GraemeL, Gravitan, GregorB, Groovenstein, Guroadrunner, Guul, Hellbus, Hellisp, Highwind, Histrion, Hyad, Hyperfusion, IcyStorm, InnocentIII, Intangir, Irdepesca572, Ixfd64, J.delanoy, Jac16888, Jafet, Jaganath, Jaycrabo, Jecowa, Jeff3000, Jjhat1, Joel7687, Joffeloff, John Reaves, Johnnaylor, Joshk, Jsbillings, Kaleb zero, Karl-Henner, KaySL, Kbh3rd, Keesiewonder, Kieff, Kiio, Leevanjackson, Lightmouse, Lionelbrits, Liontamer, Luckrider7, Makelelecba, Maraimo, MassKnowledgeLearner, Matey, Mattscool01, MegamanX64, Michael Daly, Michaelas10, Migpi, Mikkow, Minghong, Morte, Msavidge, MukiEX, Mvas, Mxn, Myscrnnm, Nebulousity, Neilc, Neodarksaver, Neutrality, Nitrodist, Nitya Dharma, Nkayesmith, Nonagonal Spider, Noodlez84, Ofbarea, Otvaltak, OuterHeaven, PS2pcGAMER, Pascal.Tesson, Password, Pathoschild, Pavel Vozenilek, Peter17, Planetary, Poolboy8, Possum, PrimeHunter, Quelloquialism, Rada, Radagast, Raysonho, Rebroad, Records, Reinis, Remember the dot, Rhobite, Ricky81682, Rjwilmsi, Rmallins, Rory096, RoyBoy, Roybb95, Ryan256, Ryk, SDBR39952, SF007, Sahkuhnder, Sam Korn, Samsara, Scepia, Scott Paeth, Sgeo, Shadowstar, Shaggorama, Shaissasas, Shello, Sietse Snel, Silvershades76, SirGrant, Snowolf, Socby19, Sparky132, Steverapaport, StoptheDatabaseState, Strait, Stunt, Tarcieri, Techdawg667, TeeEmCee, Tempshill, The Anome, The Fat Guy, Thomasda, Thumperward, Timwi, Tintazul, Tizio, Tommstein, Tumpy119, VAcharon, ViveCulture, Voyagerfan5761, Wapcaplet, WhosAsking, WiccaIrish, WipEout!, Wrightbus, Ww.ellis, Wwoods, Xuenay, Yarmo81, Youlikeyams?, ZachPruckowski, Zagen30, Zodon, Zomic13, Ásgeir IV., Александр Мотин, 374 anonymous edits

Classical mechanics *Source*: http://en.wikipedia.org/windex.php?oldid=295785794 *Contributors*: 192.169.41.xxx, 212.153.190.xxx, 62.0.98.xxx, APH, Adi Avidor, Amareto2, Amd628, Andeasling, AndrewDressel, Anjor, Antandrus, Anthony, Ap, Aravindet, AstroNomer, AtholM, Austin Maxwell, AxelBoldt, Barticus88, Bassbonerocks, Batmanand, Bcrowell, Beland, Berland, Bhadani, Bobo192, Bogdangiusca, Brews ohare, Brian0918, BrokenSegue, Bunnyhop11, C quest000, CUSENZA Mario, CYD, Camembert, Carl T, Cethegus, Charles Matthews, Cholewa, Chrylis, Complexica, CompuChip, Conversion script, Costela, Covington, CyrilThePig4, Dan Gardner, Dandrake, David R. Ingham, Db099221, DeepBlueDiamond, Deeptrivia, Denny, Dfalcantara, DirkvdM, Djr32, Dodiad, Dr. Sunglasses, Drini, Drostie, Duae Quartunciae, Ekotkie, Enormousdude, Euneirophrenia, Evil Monkey, F3meyer, FT2, Farid2053, FlorianMarquardt, Foxjwill, FreplySpang, Frédérick Lacasse, Gauss, Gene Nygaard, Giftlite, Glenn, Gnf1, Graham87, Grahamp, Gulmammad, Gwernol, Hadal, Haham hanuka, Hairy Dude, Harald88, Headbomb, Hqb, Hydrogen Iodide, Icairns, Imusade, Isis, JabberWok, Jagged 85, James086, Jayvdb, Jeepien, Jeff3000, JerrySteal, JoeSmack, Johann Wolfgang, Jomsborg, Jpc4031, Juliancolton, Kahriman, Karol Langner, Leapfrog314, Linas, Ling.Nut, Lir, Lisatwo, Logicus, Loodog, Looxix, Lrrasd, LucaB, Lupinoid, MBisanz, MarSch, MarcusMaximus, Matthew Auger, Mav, Mayooranathan, Mayz, Memayer, Michael Hardy, Mlessard, Mobius, Moink, MrXow, Muijz, Neparis, Nigelj, Nigholith, Njaelkies Lea, Nuno Tavares, Oleg Alexandrov, Orionix, Pafcu, Paolo.dL, Papadopc, Patrick, Peterlin, Phil Boswell, Phlebas, PhySusie, Phys, Pieter Kuiper, Pizza Puzzle, Pizza1512, PlatinumX, Ptpare, Quadell, Qwertyus, RE, RG2, RSido, RadicalBender, Ragesoss, RandomWalk, Raul654, Reddi, Retodon8, RexNL, Roadrunner, Rohan Ghatak, Rossami, Rracecarr, Rsm99833, Rushbie, Ruud Koot, Ryeterrell, SMP, Saeed.Veradi, Salmar, Sankalpdravid, Sanpaz, Seidenstud, Sharkface217, Silly rabbit, SirFozzie, Sketch051, SpookyMulder, Srleffler, Ssiruuk25, Stevenj, Sun King, Sure kr06, TakuyaMurata, Tarquin, Template namespace initialisation script, The Original Wildbear, Thechamelon, Thou shalt not have any gods before Willy on Wheels, Tim Starling, TimBentley, Timothyarnold85, Tom Atwood, Tom Lougheed, Tom harrison, Tomasz Prochownik, Tripodian, Truthnlove, Ulcph, Vgy7ujm, Windrixx, Wolfkeeper, Woodstone, Wwoods, XJamRastafire, XaosBits, Xunex, Yill577, Yurik, Zfr, 百家姓之四, 206 anonymous edits

Newton's laws of motion *Source*: http://en.wikipedia.org/windex.php?oldid=296067593 *Contributors*: 21655, 5IIHoova, 963kickemall, @pple, A More Perfect Onion, A-Hrafn, AJKING182, Abdullah Köroğlu, AceMyth, Adambro, Addshore, AdjustShift, Aelyn, Af648, Agge1000, Ahoerstemeier, Aitias, Aksi great, Aktsu, Al pope, Alana Shirley, Alansohn, Aleksas, Alexfusco5, AlexiusHoratius, Alexjohnc3, Alfredsimpson, Ambassador Quan, Anaraug, Ancheta Wis, AndonicO, Andres, Andrevan, AndrewDressel, Angelus Delapsus, Animum, AnonGuy, Anskas, Antandrus, Apis O-tang, Arakunem, Aremith, ArglebargleIV, Armaetin, Army1987, Arniep, Ashman512, AstroNomer, AtticusX, Avatar 06349, Avono, AxelBoldt, Axy, AzaToth, Aïki, B44H, Babygene52, BanditBubbles, Bartledan, Bcrowell, Bdesham, Bdodo1992, Ben pcc, BenFrantzDale, Bender2k14, Berl95, BertSen, Betterusername, Betterworld, Bh3u4m, Biopresto, Bobo192, Bogey97, Bomac, Bonadea, Bongwarrior, BorgQueen, BrettAllen, Brews ohare, Brianjd, Bronzie, Bth, Burntmonkey5, C'est moi, CART fan, CWii, Caiaffa, Calvin 1998, Can't sleep, clown will eat me, CanOfWorms, Canadian-Bacon, CanadianLinuxUser, Candy12324, Capricorn42, Captain Wikify, Captain-tucker, CardinalDan, Carewolf, Cb77305, Cbgermany, Celebere, Cethegus, Chairmclee, Chase me ladies, I'm the Cavalry, Chaseyoung1500, Chcknwnm, Cholmes75, Chris the speller, Cimex, Cisca Harrison, Citicat, Clayboskie, Closedmouth, Connelly, Conrad.Irwin, Conversion script, Cool3, CoyoteG, Cpastern, Crazycomputers, Creektheleftcheeksneak, Cress Arvein, Crowsnest, Cryptic, Crypticfortune, Crystallina, Cuckooman4, Curps, D, D-Notice, DJ Clayworth, Da monster under your bed, Dallas84, Daniel Case, Dark Dragon Master1337, Darkmaster333, Dave6, David.hillshafer, DavidCary, DavidRF, Dbfirs, DeadEyeArrow, Decltype, Decumanus, Deor, DerHexer, Dfrg.msc, Digmaster, Dilip rajeev, Djr32, Djsolie, Dkasak, Docu, Dodiad, Dodo von den Bergen, Dogah, Donarreiskoffer, DoubleBlue, Dougluce, Dreadstar, Dspradau, Durtysouthgurl, Dycedarg, Długosz, EJF, ERcheck, EconoPhysicist, Ed-dg, Edgar181, Eeekster, Ehabkost, El C, ElTchanggo, Eliz81, Elm-39, Elmer Clark, Enormousdude, Epbr123, Eric119, EricV89, Erik9, Euchiasmus, Everyking, Explicit, Ezelenyv, Fairchild-Republic, Fangfufu, Faradayplank, Farosdaughter, Farquaadhnchmn, Feezo, Fetofs, Fil21, Filll, Firebat08, Flame1009, FlamingSilmaril, Flewis, Flowerpotman, Foxtrotman, Fram, Frankenpuppy, Frdayeen, Fredrik, FreplySpang, Fvw, Gabrielleitao, Gandalf61, Gary2863, Gene Nygaard, Gerbrant, Gerhard.Brunthaler, Giftlite, Gilliam, GnuDoyng, Gogobera, Graham77, Graham87, Grandpsykick, Gurch, Hadal, Haham hanuka, Hakufu Sonsaku, Half-Blood Auror, HalfShadow, Hamsterlopithecus, HappyCamper, Hayabusa future, HazeNZ, Hbent, Headbomb, Hellojoshhowareyou, Heron, Heyyal77, Homerguy, Hongooi, HorsePunchKid, Hu12, Hulten, Hwefhasvs, Hydrogen Iodide, Ian13, Icairns, Icewedge, Idk my bff jill, Idont Havaname, Ihope127, Ilyushka88, Imgoș, Imlost20, Iridescent, Isaac Dupree, IsaacNewton17, Isis, Ixfd64, J Di, J.Wolfe@unsw.edu.au, J.delanoy, JCSantos, JForget, JRSpriggs, Jack1002009, Jagged 85, James R. Ward, JamesM123, JamieS93, Jason Quinn, Jcrook1987, Jeff G., Jfkdklsjf, Jgmakin, Jhud89, Jj137, JodyB, JoeBlogsDord, JoeSmack, John254, JohnCD, Johnflux, Jok2000, Joke dst, Jon Cates, Jorgenumata, Josh Parris, Jp347, Jshane04, Jzenksta, KChiu7, Kabton14, KaiMartin, Kainino, Kaishershatner, Kalathalan, Karada, Karol Langner, KathrynLybarger, Keenan Pepper, Keilana, Kerotan, Kevinsam, Kid A, Killiondude, Kimse, Kingpin13, Kipton, Kjkl, Kliao93, Kman229, Knakts, KnowledgeOfSelf, Knutux, Kool11, Korandder, Kornfan71, Krazyklink, Krea, Kribbeh, Kryic83, Kukini, Kuru, Kyle Barbour, Kzollman, LAX, La Pianista, LaFoiblesse, Lahiru k, Laky68, Laurapr, Laurascudder, Law200000, LeaveSleaves, Lectonar, Lee J Haywood, Leo44, LiDaobing, Light current, LightAnkh, Lightmouse, Lilchicklet007, Ling.Nut, Lir, Lmhjulian, LonelyBeacon, Loom91, Looxix, Lordgilman, Lpug21, Lseixas, Luna Santin, Lunchscale, M1ss1ontomars2k4, MASQUERAID, MER-C, MSGJ, Maaru, Magnus.de, Make91, Malerin, Mandarax, Mankar Camoran, MapsMan, MarSch, MarcusMaximus, Marek69, MarkSweep, Marvinfreeman, Mastercampbell, Matt 888, Matt Deres, Matthew Stannard, Matusz, Maxim, Mazca, Mboverload, McVities, Mchhabria, Mct mht, Mecanismo, Meisterkoch, Mentifisto, Michael Hardy, MichaelBillington, Mikenorton, Mindmatrix, Minish man, Misza13, Moondyne, Mr Poo, Mr. Wheely Guy, Mrboh, Mrdempsey, Muhends, Mwtoews, Mxn, Mysid, NCurse, Nabla, Nancy, Natalie Erin, Natl1, Nbauman, Neurolysis, Neverquick, NewEnglandYankee, Neyshan, Nhandler, Nicholasnice, Nicholeeeeo, Nickpowerz, Nicop (Usurp), Nihiltres, Nilfanion, NinjaKid, Nishkid64, Nja247, Nk, Nneonneo, No Guru, Noah Salzman, Nommonomanac, Notalex, Novakyu, Obli, Obvioustrollisobvious, Ocvailes, Oda Mari, Ohnoitsjamie, Oleg Alexandrov, Oolongy, Opelio, OrbitOne, Orfen, Oxymoron83, P3d0, PMDrive1061, PaddyLeahy, Paolo.dL, Passamaquoddy boi, Pasteman, Patrick, Pedro, Pele' boy, Peregrine Fisher, Perfgeek, Persian Poet Gal, Peterlin, Pevarnj, Pewwer42, Pfalstad, Philip Trueman, Phinnaeus, PhySusie, Pizza Puzzle, Pkbharti, Pleasantville, Plrk, Pmetzger, Polluxian, Pomona17, Possum, Prashanthns, Proofreader77, Proud Muslim, Puchiko, Puneetbahri 82, Quantumobserver, R'n'B, R. Baley, RFerreira, RG2, RHB100, RJaguar3, RainbowOfLight, Rama's Arrow, RandomXYZb, Raven in Orbit, RayAYang, Raziaex, Red Thunder, Redfarmer, Res2216firestar, RexNL, Richard75, Ricky81682, Rje, Rnt20, Robdurbar, Robin Patterson, Robinh, Rogper, Rokfaith, Rolo Tamasi, RoyBoy, Rracecarr, Rudjek, S Schaffter, S3000, SFC9394, Salsa Shark, Salt Yeung, Sandahl, Saros136, Sarregouset, Schipperke22, Science4sail, ScienceApologist, Scigatt, Sciurinæ, Scizor55, Sean William, Sergay, Sethmiester, Shanes, Sharonlees, Shastra, Shaverc, Shearyears394, Sheliak, Shirulashem, Shizhao, Shockkorea, Shoy, Sibusiso Mabaso, Siddhant, Silly rabbit, Simetrical, Simon123, SimonP, Sionus, SirGrant, Skier Dude, Sl, Slakr, Smack, Smokizzy, Snowmanradio, SoWhy, Sodium, Sokane, Solitude, Sonia62585, Sowelilitokiemu, Spaceman85, Spaceneil8, Specter01010, SpeedyGonsales, SperryTS, Spinningspark, Splash, SpookyMulder, StaticGull, Staxringold, Stephenb, SteveBaker, Stevertigo, Stewartadcock, StradivariusTV, Stuhacking, Stui, Stwalkerster, Subheight640, Sunilbajpai, Susanwangrules, Sverdrup, Sword, Synchronism, TStein, Tagishsimon, Tameradly, Taneli HUUSKONEN, Tcncv, Teressa Keiner, ThaddeusB, The Rambling Man, The wub, Thingg, Thljcl, Tiddly Tom, Tiptoety, Tnxman307, Tohd8BohaithuGh1, Tom Lougheed, Tombomp, Toxic Spade, Transcendence, Travelbird, Treelo, Trevor MacInnis, Trevor Marron, Truthnlove, Tseulik, Tyomitch, UberScienceNerd, Ubiq, Uksam88, Unmerklich, Urhixidur, Urmammasfat, Valandil211, VashiDonsk, Venfranc, Versus22, Vgm3985, VolatileChemical, Voyajer, Waggers, WarthogDemon, Wenli, WereSpielChequers, Wesino, Wienwei, Wiki alf, Wikieditor06, Wimt, Wolfkeeper, Wsvlqc, Wwoods, X42bn6, Xaven, Xxanthippe, Xy7, Xykon, Yamamoto Ichiro, YellowMonkey, Yevgeny Kats, Yossiea, Yoyoyow, Yuksing, Yurigerhard, Yurik, Yuyudevil, Zadeez, Zenzic, Zephyr9, Zginder, Zsinj, Zzyzx11, □□□□□□□□□, 1862 anonymous edits

Analytical dynamics *Source*: http://en.wikipedia.org/windex.php?oldid=292074711 *Contributors*: Brews ohare, Djr32, Ideal gas equation, RHB100, Rich Farmbrough, Sanpaz, 1 anonymous edits

Molecular dynamics *Source*: http://en.wikipedia.org/windex.php?oldid=295309203 *Contributors*: Agilemolecule, Alex.g, Amire80, Ammatsun, Anthracene, Anxdo, ApJilly, Astavats, Ayucat, Bbullot, Bduke, BenFrantzDale, Bubba hotep, Chris the speller, Cortonin, Cwassman, DMacks, DRider, Dacb, DeadEyeArrow, Demus Wiesbaden, Dicklyon, Dietmar.paschek, DragonflySixtyseven, Drswenson, Ebuchol, Ehdr, Gentgeen, Giftlite, Huckit, Itamblyn, Itub, JWSchmidt, Jerome Charles Potts, Jorgenumata, Jugander, Kaihsu, Karol Langner, Katherine Folsom, Kennylam, Kevyn, Kjaergaard, Knordlun, Laurent1979, Lexor, LiDaobing, Linas, Lomenoldur, Ludx, Maduixa, Marchuta, Marx Gomes, Mateusz Galuszka, Mattopia, Md Arshad Iqbal, Mihoopes, Mr Marie Weaver, Msuzen, Nicolasbock, Oiramrasec, Opabinia regalis, Ossi, P99am, Paul.raymond.brenner, Pedrito, Pelister, PhCOOH, Pksach, PrometheusX303, Raviwiki4, Rob Hooft, Roo1812, Sandycx, Shura58, Smoe, Smremde, Stewartadcock, Sudiarta, TStein, Themfromspace, Thorwald, Utcursch, Van helsing, Whanrott, Wikimcmd, Wittgenstein77, Wiz9999, Xavier andrade, Yrtgm, 200 anonymous edits

CHARMM *Source*: http://en.wikipedia.org/windex.php?oldid=296182826 *Contributors*: Betacommand, Btmiller, Closedmouth, DMacks, Dmb000006, Dysprosia, Headbomb, ISGTW, Itub, Jag123, Jedgold, Joerg Kurt Wegner, Johnpacklambert, Jorgenumata, Kaihsu, Karol Langner, LizGere, Mark Oakley, Mndoci, P99am, PhCOOH, Rich Farmbrough, Shura58, Taw, ThomasK, Thorwald, Trevyn, Vicki Rosenzweig, Vincent kraeutler, 23 anonymous edits

Statistical mechanics *Source*: http://en.wikipedia.org/windex.php?oldid=294155089 *Contributors*: APH, Abtract, Aleksas, Alison, Ancheta Wis, Andries, Anoko moonlight, Ap, Boardhead, Bogdangiusca, Brews ohare, Brianjd, Bryan Derksen, BryanD, Charles Matthews, Charvest, Chrisch, Complexica, Conversion script, Cordell, DMTagatac, Dancter, Daniel5127, Davennmarr, David Woolley, Den fjättrade ankan, DiceDiceBaby, Djr32, Djus,

Doprendek, Edgar181, Edkarpov, Edsanville, Elwikipedista, Eman, Fephisto, Frokor, G716, Gail, GangofOne, Giftlite, Gurch, HappyCamper, Ht686rg90, Isopropyl, IvanLanin, JabberWok, Jheald, John Palkovic, Jorgenumata, Karl-Henner, Karol Langner, Kzollman, Lantonov, LeadSongDog, Linas, Linuxlad, Locke9k, Loupeter, Lyonspen, MK8, Mary blackwell, Mct mht, Michael Hardy, Michael L. Kaufman, Miguel, Mikez, Mlys, Monedula, Moondarkx, Mpatel, Mxn, Nitya Dharma, Nnh, P99am, PAR, Patrick, Peabeejay, Perelaar, Peterlin, Phudga, Phys, PhysPhD, Politepunk, Radagast83, RandomP, Rjwilmsi, Ryanmcdaniel, SDC, Sadi Carnot, Samkung, Scorwin, Sheliak, SimpsonDG, Steve Omohundro, StewartMH, StradivariusTV, Template namespace initialisation script, That Guy, From That Show!, The Anome, The Cunctator, The Original Wildbear, The.orpheus, Thingg, ThorinMuglindir, Tim Starling, TravisAF, Truthnlove, Van helsing, Vql, Wavelength, Weiguxp, Wiki me, Xp54321, Yevgeny Kats, Yill577, 114 anonymous edits

Statistical field theory *Source*: http://en.wikipedia.org/windex.php?oldid=251854946 *Contributors*: Dolphin238, G716, Laurascudder, Maliz, Oleg Alexandrov, R.e.b., Stevvers, 16 anonymous edits

Computational chemistry *Source*: http://en.wikipedia.org/windex.php?oldid=294143278 *Contributors*: 194.200.130.xxx, 4lex, 66.122.54.xxx, APH, AcidFlask, Agilemolecule, Ahoerstemeier, Alexandrov, Altenmann, Andyras, Annabel, Arteum, Ashujo, BRG, Bduke, BenFrantzDale, CYD, Cdr Harris Fan Club, Charles Matthews, Chmrjg, ConceptExp, Conversion script, Cyan, DV8 2XL, Edsanville, El C, Epbr123, Feitclub, Fuzheado, Gentgeen, Ghutchis, Giftlite, Giggy, Glloq, Grimlock, Gseryakov, Gurch, HRV, HappyCamper, Harold f, Hlwoodcock, Hugo-cs, Huttite, Iridium77, Isidore, Isilanes, Itub, JJL, James 007, Jantony420, JerrySteal, Jitse Niesen, Jocelyne Heys-Gerard, JustMyself, Kaihsu, Kaiserkarl13, Karol Langner, Kku, Kmarkey, Knights who say ni, Larry laptop, Lawrence O'Neil, Lexor, Lfh, LiDaobing, LuchoX, Marcoacostareyes, Marj Tiefert, Maurreen, Mayumashu, Mboverload, Michael Devore, Moink, Mxn, NNemec, NuclearWinner, Od Mishehu, Okedem, Oleg Alexandrov, Oneboy, OrgasGirl, P.wormer, P99am, Paul August, PhCOOH, Poszwa, ProteusCoop, Quanda, Ralf Schmelter, Rjwilmsi, Ronnotel, Roswell native, Rpetrenko, Ruud Koot, Sadi Carnot, Sam Hocevar, Samuel Grant, Sbo, SeventyThree, Shyamal, Spike Wilbury, Stewartadcock, Stone, Swheele2, Themfromspace, TimBentley, Timwi, Tomturner, Tomtzigt, Van helsing, Vb, Vina, Voorlandt, Whiner01, ~K, 154 anonymous edits

Mathematical chemistry *Source*: http://en.wikipedia.org/windex.php?oldid=295232989 *Contributors*: Altenmann, Baz.77.243.99.32, DMacks, Giftlite, Hamleteer, Itub, Linas, Maxal, Mpatel, Niksab, Remuel, Tim32, 10 anonymous edits

Monte Carlo method *Source*: http://en.wikipedia.org/windex.php?oldid=295834903 *Contributors*: *drew, ABCD, Aardvark92, Adfred123, Aferistas, Agilemolecule, Akanksh, Alanbly, Albmont, AlexBIOSS, AlexandreCam, AlfredR, Alliance09, Altenmann, Andrea Parri, Andreas Kaufmann, Angelbo, Aniu, Apanag, Aspuru, Atlant, Avalcarce, Aznrocket, BAxelrod, BConleyEEPE, Banano03, Banus, Bduke, BenFrantzDale, BenTrotsky, Bender235, Bensaccount, BillGosset, Bkell, Blotwell, Bmaddy, Bobo192, Boffob, Boredzo, Broquaint, Btyner, CRGreathouse, Caiaffa, Charles Matthews, ChicagoActuary, Cibergili, Cm the p, Colonies Chris, Coneslayer, Cretog8, Criter, Cybercobra, Cython1, DMG413, Damistmu, Ddcampayo, Ddxc, Digemedi, Ds53, Duck ears, Duncharris, Dylanwhs, ERosa, EldKatt, Elpincha, Elwikipedista, Eudaemonic3, Ezrakilty, Fastfission, Fintor, Flammifer, Frozen fish, Furrykef, G716, Giftlite, Gilliam, Goudzovski, GraemeL, GrayCalhoun, Greenyoda, Grestrepo, Gtrmp, Gökhan, Hanksname, Hawaiian717, Hokanomono, Hu12, Hubbardaie, ILikeThings, IanOsgood, Inrad, Itub, Jackal irl, Janpedia, JavaManAz, Jeffq, Jitse Niesen, Joey0084, JohnOwens, Jorgenumata, Jsarratt, Jugander, Jérôme, K.lee, KSmrq, KaHa242, Karol Langner, Kenmckinley, Kimys, Knordlun, Kroese, Kummi, Kuru, Lambyte, LeoTrottier, Levin, Lexor, LoveMonkey, Malatesta, Male1979, ManchotPi, Marcofalcioni, Martinp, Masatran, Mathcount, MaxHD, Maxentrope, Maylene, Melcombe, Michael Hardy, Mikael V, Misha Stepanov, Mnath, Moink, Mtford, Nagasaka, Nanshu, Narayanese, Nelson50, Nosophorus, Nsaa, Nuno Tavares, Nvartaniucsd, Ohnoitsjamie, Oli Filth, Oneboy, Orderud, OrgasGirl, P99am, Paul August, PaulxSA, Pbroks13, Pcb21, Pete.Hurd, PeterBoun, Pgreenfinch, Philopp, Pibwl, Pinguin.tk, PlantTrees, Pne, Popsracer, Poupoune5, Qadro, Quantumelfmage, Quentar, Qxz, RWillwerth, Ramin Nakisa, Redgolpe, Renesis, Richie Rocks, Rinconsoleao, Rjmccall, Ronnotel, Rs2, SKelly1313, Sam Korn, Samratvishaljain, Sergio.ballestrero, Shreevatsa, Snoyes, Somewherepurple, Spellmaster, Splash6, SpuriousQ, Stefanez, Stefanomione, StewartMH, Stimpy, Storm Rider, Superninja, Tarantola, Taxman, Tayste, Tesi1700, Theron110, Thirteenity, Tiger Khan, Tim Starling, Tom harrison, Tooksteps, Trebor, Twooars, Urdutext, Vipuser, VoseSoftware, Wile E. Heresiarch, Yoderj, Zarniwoot, Zoicon5, Zr40, Zuidervled, 308 anonymous edits

Quantum Monte Carlo *Source*: http://en.wikipedia.org/windex.php?oldid=295037474 *Contributors*: Acipsen, Amyoungil, Bci2, Conscious, Henry Delforn, Isilanes, Jheriko, Karol Langner, Lucaskw, Mdt26a, Melcombe, Michael Hardy, NNemec, Pablomme, Paulcardan, Rbonvall, Rich Farmbrough, Rjwilmsi, Sagaciousuk, Supersion77, TestPilot, Trigger hippie77, UkPaolo, Veinor, Vgy7ujm, Vyznev Xnebara, WilliamDParker, Wirawan0, 38 anonymous edits

Dynamics of Markovian particles *Source*: http://en.wikipedia.org/windex.php?oldid=275949632 *Contributors*: Bergner, Czolgolz, ERobson, Fabrictramp, Icairns, JPG-GR, Malcolma, Mbell, Oleg Alexandrov, ShakespeareFan00, 14 anonymous edits

Metabolic network *Source*: http://en.wikipedia.org/windex.php?oldid=231870375 *Contributors*: Blastwizard, Ceyockey, Oleginger, PDH, TheParanoidOne, TimVickers, Zephyris, 2 anonymous edits

Topological dynamics *Source*: http://en.wikipedia.org/windex.php?oldid=285184911 *Contributors*: Arcfrk, Charvest, Michael Hardy, SiamakT

Protein folding *Source*: http://en.wikipedia.org/windex.php?oldid=295343721 *Contributors*: 168..., 5beta5, Adriferr, Agilemolecule, Akane700, Andraaide, Arcadian, Banus, Barticus88, Bci2, Bendzh, Bfinn, Bioinfo177, Biophys, Biophysik, Blainster, Blooooo, Brianga, Bryan Derksen, Cacycle, Calvero JP, Cathalgarvey, Cburnett, ChicXulub, Clicketyclack, Computor, Cyberman, Czhangrice, D. Recorder, DannyWilde, Davepntr, DennisDaniels, Dhatz, Donarreiskoffer, Dwmyers, ESkog, Eequor, Erencexor, Erwinser, Fawzin, Fuzheado, Gcrossan, Gowantervo, Grimlock, Harley peters, Herd of Swine, Hooplehead, Intangir, Ixfd64, JJ TKOB, Jacobsman, Jammedshut, JeramieHicks, Katherine Folsom, Kevyn, Kierano, Kjaergaard, Konstantin, Kostmo, Leptictidium, Lexor, Lfh, LiDaobing, Lir, Lostart, Lucaaah, M stone, Macintosh10000, Madeleine Price Ball, Magnus Manske, Malcolm Farmer, Mark Renier, Michael Hardy, Miguel Andrade, Minghong, MockAE, Movado73, Movalley, Myscrnnm, Netesq, Opabinia regalis, Otvaltak, P99am, Piotrus, Ptr123, RDBrown, Rebroad, Richyfp, Rjwilmsi, RoyBoy, RunninRiot, Samrat Mukhopadhyay, Sharkman, Shmedia, SirHaddock, Smelialichu, Snowmanradio, Splette, SpuriousQ, Stepa, Sunny17152, Taw, Terrace4, TestPilot, The Anome, The Great Cook, Tijmz, TimVickers, Tomixdf, Tommstein, Toytoy, Trenchcoatjedi, Trusilver, V8rik, Waerloeg, WillowW, WorldDownInFire, Xnuala, Xy7, Yves-henri, Zoicon5, ראיל, 147 anonymous edits

Protein-protein interaction *Source*: http://en.wikipedia.org/windex.php?oldid=293638290 *Contributors*: 56869kltaylor, 7bd1, A wandering 1, Alboyle, Apfelsine, Ashcroft, Bci2, Biophys, Clicketyclack, Cpichardo, D-rew, DarkSaber2k, Delldot, Djstates, Dsome, FreeKill, Giftlite, GracelinTina, Hendrik Fuß, Hotheartdog, Jeandré du Toit, Jkbioinfo, Jkwaran, Jn3v16, Jongbhak, Keesiewonder, Kkmurray, Kuheli, Kyawtun, Lafw, Lemchesvej, Lenticel, Longhair, Meb025, Michael Hardy, MichaelMcGuffin, Miguel Andrade, NickelShoe, Ninjagecko, Nnh, Rajah, Reb42, Riana, Ronz, Seans Potato Business, Snowolf, TheParanoidOne, Thorwald, Uthbrian, Victor D, Wenzelr, Whosasking, Wintrag, 67 anonymous edits

DNA Dynamics *Source*: http://en.wikipedia.org/windex.php?oldid=296104632 *Contributors*: Auntof6, Bci2, Cander0000, Chris the speller, CommonsDelinker, Ironholds, Potatoswatter

DNA nanotechnology *Source*: http://en.wikipedia.org/windex.php?oldid=294763628 *Contributors*: 0x38I9J*, Alnokta, Amaling, Anthonydelaware, Antony-22, Cyfal, Epbr123, Giftlite, Gioto, Pwkr, ShawnDouglas, Thorwald, Tolosthemagician, ZayZayEM, 30 anonymous edits

Molecular self-assembly *Source*: http://en.wikipedia.org/windex.php?oldid=295612777 *Contributors*: Antony-22, Dirac66, LithoGuy, M stone, Materialscientist, Maurog, Mdd, Nanotrix, Netmonger, Northfox, Rudick.JG, Satish.murthy, TestPilot, 21 anonymous edits

Cell signaling *Source*: http://en.wikipedia.org/windex.php?oldid=294194237 *Contributors*: A-Day, Amenzix, Annekcm, Apfelsine, Arcadian, B9 hummingbird hovering, Biochemza, Clicketyclack, Computerjoe, Dlohcierekim, DrTLesterThomas, Drphilharmonic, Edgar181, Fannaba, Fritzpoll, Hoverflysmiles, JWSchmidt, K.murphy, Lenticel, Letranova, Lkathmann, Lunska, Mandarax, Mdd, Mikael Häggström, Nanobri, Nemenman, Neutrality, Nihiltres, OldakQuill, PaddyM, Physicistjedi, Popo le Chien, Pozzie, Radagast83, Rjwilmsi, Sardaukar Blackfang, Seans Potato Business, Shell Kinney, Squidonius, Tabletop, Tameeria, Tikiwont, Vary, Wisdom89, Wolfrock, Woohookitty, Ynaztiw, 49 anonymous edits

Molecular evolution *Source*: http://en.wikipedia.org/windex.php?oldid=294465595 *Contributors*: 10outof10die, 168..., A.bit, AdamRetchless, Aranae, Aunt Entropy, AxelBoldt, Ben Tillman, Borgx, Bornhj, Debresser, Duncharris, Emw2012, Etxrge, Eugene van der Pijll, Ewen, GSlicer, Gaius Cornelius, GeoMor, Kosigrim, Lexor, Lindosland, M stone, MER-C, Marooned Morlock, Mcw1139, Neutrality, Nonsuch, Northfox, Notreallydavid, OnBeyondZebrax, Owenman, PDH, PhDP, Ragesoss, Rigadoun, Ryulong, Sadi Carnot, Samsara, Seglea, Shyamal, Steinsky, Stirling Newberry, StormBlade, Swpb, Template namespace initialisation script, That Guy, From That Show!, The Anome, Theuser, Thue, Timwi, Vsmith, Wavesmikey, Whatiguana, 52 anonymous edits

Molecular phylogenetics *Source*: http://en.wikipedia.org/windex.php?oldid=295503179 *Contributors*: 168..., Aranae, Brya, Carlosp420, Dysmorodrepanis, EdJohnston, Eugene van der Pijll, Florentino floro, Igodard, JForget, JoJan, Joseph Solis in Australia, Kingdon, Lexor, M1ss1ontomars2k4, Mario1952, Marj Tiefert, Onco p53, Otherlleft, PierreAbbat, Radagast83, Ragesoss, RoRo en wiki, Samsara, Shyamal, Sinneed, Stemonitis, StephenWeber, Styrofoam1994, 20 anonymous edits

Computational phylogenetics *Source*: http://en.wikipedia.org/windex.php?oldid=281615251 *Contributors*: Aranae, Dysmorodrepanis, Fred Hsu, Hongooi, J. Spencer, JohnnyHom, Lambiam, Leonard G., MrDolomite, Noleander, Opabinia regalis, Ramesan, Thorwald, Whatiguana, Wild8oar, Wzhao553, 9 anonymous edits

Image Sources, Licenses and Contributors

Image:peptide angles.png *Source*: http://en.wikipedia.org/windex.php?title=File:Peptide_angles.png *License*: unknown *Contributors*: User:Bensaccount

Image:Hardware-accelerated-molecular-modeling.png *Source*: http://en.wikipedia.org/windex.php?title=File:Hardware-accelerated-molecular-modeling.png *License*: Public Domain *Contributors*: User:P99am

Image:LCAOwater1.JPG *Source*: http://en.wikipedia.org/windex.php?title=File:LCAOwater1.JPG *License*: Public Domain *Contributors*: ChemGrrl, Keenan Pepper

Image:SimpleMOdiagram.png *Source*: http://en.wikipedia.org/windex.php?title=File:SimpleMOdiagram.png *License*: unknown *Contributors*: V8rik

Image:jmol1.png *Source*: http://en.wikipedia.org/windex.php?title=File:Jmol1.png *License*: unknown *Contributors*: Peter Murray-Rust

Image:Hemagglutinin molecule.png *Source*: http://en.wikipedia.org/windex.php?title=File:Hemagglutinin_molecule.png *License*: GNU Free Documentation License *Contributors*: U.S. National Institutes of Health.

Image:FormicAcid.pdb.png *Source*: http://en.wikipedia.org/windex.php?title=File:FormicAcid.pdb.png *License*: Public Domain *Contributors*: Csörföly D, Cwbm (commons)

Image:Isosurface on molecule.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Isosurface_on_molecule.jpg *License*: unknown *Contributors*: StoatBringer

Image:porin.qutemol.dl.png *Source*: http://en.wikipedia.org/windex.php?title=File:Porin.qutemol.dl.png *License*: unknown *Contributors*: ALoopingIcon

Image:porin.qutemol.ao.png *Source*: http://en.wikipedia.org/windex.php?title=File:Porin.qutemol.ao.png *License*: unknown *Contributors*: ALoopingIcon

Image:JmolStick.PNG *Source*: http://en.wikipedia.org/windex.php?title=File:JmolStick.PNG *License*: unknown *Contributors*: Peter Murray-Rust

Image:PEF comparison.png *Source*: http://en.wikipedia.org/windex.php?title=File:PEF_comparison.png *License*: unknown *Contributors*: User:Edboas

Image:Sarfus.DNABiochip.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Sarfus.DNABiochip.jpg *License*: unknown *Contributors*: Nanolane

Image:Spinning DNA.gif *Source*: http://en.wikipedia.org/windex.php?title=File:Spinning_DNA.gif *License*: Public Domain *Contributors*: USDA

File:Methanol.pdb.png *Source*: http://en.wikipedia.org/windex.php?title=File:Methanol.pdb.png *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: ALoopingIcon, Benjah-bmm27

File:DNA-fragment-3D-vdW.png *Source*: http://en.wikipedia.org/windex.php?title=File:DNA-fragment-3D-vdW.png *License*: Public Domain *Contributors*: Benjah-bmm27

File:Simple harmonic oscillator.gif *Source*: http://en.wikipedia.org/windex.php?title=File:Simple_harmonic_oscillator.gif *License*: Public Domain *Contributors*: User:Oleg Alexandrov

File:DNA chemical structure.svg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_chemical_structure.svg *License*: unknown *Contributors*: Madprime, Wickey, 1 anonymous edits

File:ADN animation.gif *Source*: http://en.wikipedia.org/windex.php?title=File:ADN_animation.gif *License*: Public Domain *Contributors*: Aushulz, Bawolff, Brian0918, Kersti Nebelsiek, Magadan, Mattes, Origamiemensch, Stevenfruitsmaak, 3 anonymous edits

File:Parallel telomere quadruple.png *Source*: http://en.wikipedia.org/windex.php?title=File:Parallel_telomere_quadruple.png *License*: unknown *Contributors*: User:Splette

File:Four-way DNA junction.gif *Source*: http://en.wikipedia.org/windex.php?title=File:Four-way_DNA_junction.gif *License*: Public Domain *Contributors*: Aushulz, Molatwork, Origamiemensch, TimVickers, 1 anonymous edits

File:DNA replication.svg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_replication.svg *License*: unknown *Contributors*: Bibi Saint-Pol, Elborgo, En rouge, Helix84, Jnpet, LadyofHats, Leafnode, M.Komorniczak, MetalGearLiquid, Nikola Smolenski, Nl74, Sentausa, WarX, 2 anonymous edits

File:ABDNAxrgpj.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:ABDNAxrgpj.jpg *License*: GNU Free Documentation License *Contributors*: I.C. Baianu et al.

File:Plos VHL.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Plos_VHL.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Akinom, Anniolek, Filip em, Thommiddleton

File:3D model hydrogen bonds in water.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:3D_model_hydrogen_bonds_in_water.jpg *License*: GNU Free Documentation License *Contributors*: User:snek01

Image:Methanol.pdb.png *Source*: http://en.wikipedia.org/windex.php?title=File:Methanol.pdb.png *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: ALoopingIcon, Benjah-bmm27

Image:DNA-(A)80-model.png *Source*: http://en.wikipedia.org/windex.php?title=File:DNA-(A)80-model.png *License*: unknown *Contributors*: User:P99am

File:Bragg diffraction.png *Source*: http://en.wikipedia.org/windex.php?title=File:Bragg_diffraction.png *License*: GNU General Public License *Contributors*: user:hadmack

File:DNA in water.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_in_water.jpg *License*: unknown *Contributors*: User:Bbkkk

File:X ray diffraction.png *Source*: http://en.wikipedia.org/windex.php?title=File:X_ray_diffraction.png *License*: unknown *Contributors*: Thomas Splettstoesser

File:X Ray Diffractometer.JPG *Source*: http://en.wikipedia.org/windex.php?title=File:X_Ray_Diffractometer.JPG *License*: GNU Free Documentation License *Contributors*: Ff02::3, Pieter Kuiper

File:SLAC detector edit1.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:SLAC_detector_edit1.jpg *License*: unknown *Contributors*: User:Mfield, User:Starwiz

File:ISIS exptal hall.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:ISIS_exptal_hall.jpg *License*: unknown *Contributors*: wurzeller

File:Dna-SNP.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Dna-SNP.svg *License*: unknown *Contributors*: User:Gringer

File:DNA Under electron microscope Image 3576B-PH.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_Under_electron_microscope_Image_3576B-PH.jpg *License*: unknown *Contributors*: Original uploader was SeanMack at en.wikipedia

File:DNA Model Crick-Watson.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_Model_Crick-Watson.jpg *License*: Public Domain *Contributors*: User:Alkivar

File:DNA labels.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_labels.jpg *License*: GNU Free Documentation License *Contributors*: User:Raul654

File:AT DNA base pair pt.svg *Source*: http://en.wikipedia.org/windex.php?title=File:AT_DNA_base_pair_pt.svg *License*: Public Domain *Contributors*: User:Lijealso

File:A-B-Z-DNA Side View.png *Source*: http://en.wikipedia.org/windex.php?title=File:A-B-Z-DNA_Side_View.png *License*: Public Domain *Contributors*: Original uploader was Thorwald at en.wikipedia

File:Museo Príncipe Felipe. ADN.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Museo_Príncipe_Felipe._ADN.jpg *License*: Creative Commons Attribution-Sharealike 2.0 *Contributors*: Fernando

File:AGCT DNA mini.png *Source*: http://en.wikipedia.org/windex.php?title=File:AGCT_DNA_mini.png *License*: unknown *Contributors*: Iquo

File:BU Bio5.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:BU_Bio5.jpg *License*: Creative Commons Attribution-Sharealike 2.0 *Contributors*: Original uploader was Elapied at fr.wikipedia

File:Circular DNA Supercoiling.png *Source*: http://en.wikipedia.org/windex.php?title=File:Circular_DNA_Supercoiling.png *License*: GNU Free Documentation License *Contributors*: Richard Wheeler (Zephyris)

File:Rosalindfranklinsjokecard.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Rosalindfranklinsjokecard.jpg *License*: unknown *Contributors*: Bci2, J.delanoy, Martyman, Nitramrekcap, Rjm at sleepers, 3 anonymous edits

Image:Rosalindfranklinsjokecard.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Rosalindfranklinsjokecard.jpg *License*: unknown *Contributors*: Bci2, J.delanoy, Martyman, Nitramrekcap, Rjm at sleepers, 3 anonymous edits

File:Genomics GTL Pictorial Program.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Genomics_GTL_Pictorial_Program.jpg *License*: Public Domain *Contributors*: Mdd

File:RNA pol.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:RNA_pol.jpg *License*: Public Domain *Contributors*: InfoCan

File:Primase 3B39.png *Source*: http://en.wikipedia.org/windex.php?title=File:Primase_3B39.png *License*: Public Domain *Contributors*: own work

File:DNA Repair.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_Repair.jpg *License*: Public Domain *Contributors*: Courtesy of Tom Ellenberger, Washington University School of Medicine in St. Louis.

File:MGMT+DNA 1T38.png *Source*: http://en.wikipedia.org/windex.php?title=File:MGMT+DNA_1T38.png *License*: Public Domain *Contributors*: own work

File:DNA damaged by carcinogenic 2-aminofluorene AF .jpg *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_damaged_by_carcinogenic_2-aminofluorene_AF_.jpg *License*: Public Domain *Contributors*: Brian E. Hingerty, Oak Ridge National Laboratory Suse Broyde, New York University Dinshaw J. Patel, Memorial Sloan Kettering Cancer Center

File:A-DNA orbit animated small.gif *Source*: http://en.wikipedia.org/windex.php?title=File:A-DNA_orbit_animated_small.gif *License*: GNU Free Documentation License *Contributors*: User:Bstlee, User:Zephyris

File:Plasmid emNL.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Plasmid_emNL.jpg *License*: GNU Free Documentation License *Contributors*: Denniss, Glenn, Rasbak

File:Chromatin chromosom.png *Source*: http://en.wikipedia.org/windex.php?title=File:Chromatin_chromosom.png *License*: Public Domain *Contributors*: User:Magnus Manske

File:Chromosome.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Chromosome.svg *License*: unknown *Contributors*: User:Dietzel65, User:Magnus Manske, User:Tryphon

File:Chr2 orang human.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Chr2_orang_human.jpg *License*: Creative Commons Attribution-Sharealike 2.5 *Contributors*: Verena Schubel, Stefan Müller, Department Biologie der Ludwig-Maximilians-Universität München.

File:3D-SIM-3 Prophase 3 color.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:3D-SIM-3_Prophase_3_color.jpg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Lothar Schermelleh

File:Chromosome2 merge.png *Source*: http://en.wikipedia.org/windex.php?title=File:Chromosome2_merge.png *License*: Public Domain *Contributors*: Original uploader was Evercat at en.wikipedia

File:Transkription Translation 01.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Transkription_Translation_01.jpg *License*: Public Domain *Contributors*: User:Kuebi

File:RibosomaleTranskriptionsEinheit.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:RibosomaleTranskriptionsEinheit.jpg *License*: GNU Free Documentation License *Contributors*: User:Merops

File:Chromosome Conformation Capture Technology.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Chromosome_Conformation_Capture_Technology.jpg *License*: Public Domain *Contributors*: User:Kangyun1985

File:Mitochondrial DNA and diseases.png *Source*: http://en.wikipedia.org/windex.php?title=File:Mitochondrial_DNA_and_diseases.png *License*: unknown *Contributors*: User:XXXL1986

File:PCR.svg *Source*: http://en.wikipedia.org/windex.php?title=File:PCR.svg *License*: unknown *Contributors*: User:Madprime

File:Pcr gel.png *Source*: http://en.wikipedia.org/windex.php?title=File:Pcr_gel.png *License*: GNU Free Documentation License *Contributors*: Habj, Ies, PatríciaR, Retama, Saperaud

File:DNA nanostructures.png *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

File:SFP discovery principle.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:SFP_discovery_principle.jpg *License*: unknown *Contributors*: User:Agbiotec

File:Cdnaarray.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Cdnaarray.jpg *License*: unknown *Contributors*: Mangapoco

File:Expression of Human Wild-Type and P239S Mutant Palladin.png *Source*: http://en.wikipedia.org/windex.php?title=File:Expression_of_Human_Wild-Type_and_P239S_Mutant_Palladin.png *License*: unknown *Contributors*: see above

File:Random genetic drift chart.png *Source*: http://en.wikipedia.org/windex.php?title=File:Random_genetic_drift_chart.png *License*: unknown *Contributors*: User:Professor marginalia

File:Co-dominance Rhododendron.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Co-dominance_Rhododendron.jpg *License*: Creative Commons Attribution 2.0 *Contributors*: Ayacop, Cillas, FlickrLickr, FlickreviewR, Horcha, Kanonkas, Kevmin, MPF, Para

File:DNA_nanostructures.png *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

File:Holliday junction coloured.png *Source*: http://en.wikipedia.org/windex.php?title=File:Holliday_junction_coloured.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

File:Holliday Junction cropped.png *Source*: http://en.wikipedia.org/windex.php?title=File:Holliday_Junction_cropped.png *License*: GNU Free Documentation License *Contributors*: Original uploader was TimVickers at en.wikipedia

File:Atomic force microscope by Zureks.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Atomic_force_microscope_by_Zureks.jpg *License*: unknown *Contributors*: User:Zureks

File:Atomic force microscope block diagram.png *Source*: http://en.wikipedia.org/windex.php?title=File:Atomic_force_microscope_block_diagram.png *License*: Public Domain *Contributors*: Original uploader was Askewmind at en.wikipedia

File:AFM view of sodium chloride.gif *Source*: http://en.wikipedia.org/windex.php?title=File:AFM_view_of_sodium_chloride.gif *License*: Public Domain *Contributors*: Courtesy of prof. Ernst Meyer, university of Basel

File:Single-Molecule-Under-Water-AFM-Tapping-Mode.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Single-Molecule-Under-Water-AFM-Tapping-Mode.jpg *License*: unknown *Contributors*: User:Yurko

File:AFMimageRoughGlass20x20.png *Source*: http://en.wikipedia.org/windex.php?title=File:AFMimageRoughGlass20x20.png *License*: Public Domain *Contributors*: Chych

File:Maldi informatics figure 6.JPG *Source*: http://en.wikipedia.org/windex.php?title=File:Maldi_informatics_figure_6.JPG *License*: Public Domain *Contributors*: Rbeavis

File:Stokes shift.png *Source*: http://en.wikipedia.org/windex.php?title=File:Stokes_shift.png *License*: unknown *Contributors*: User:Mykhal

File:CARS Scheme.svg *Source*: http://en.wikipedia.org/windex.php?title=File:CARS_Scheme.svg *License*: unknown *Contributors*: Onno Gabriel

File:HyperspectralCube.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:HyperspectralCube.jpg *License*: Public Domain *Contributors*: Dr. Nicholas M. Short, Sr.

File:MultispectralComparedToHyperspectral.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:MultispectralComparedToHyperspectral.jpg *License*: Public Domain *Contributors*: Dr. Nicholas M. Short, Sr.

File:Confocalprinciple.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Confocalprinciple.svg *License*: GNU Free Documentation License *Contributors*: Danh

File:3D-SIM-1 NPC Confocal vs 3D-SIM detail.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:3D-SIM-1_NPC_Confocal_vs_3D-SIM_detail.jpg *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Changes in layout by the uploader. Only the creator of the original (Lothar Schermelleh) should be credited.

File:Tirfm.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Tirfm.svg *License*: Public Domain *Contributors*: Dawid Kulik

File:Inverted microscope.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Inverted_microscope.jpg *License*: unknown *Contributors*: Nuno Nogueira (Nmnogueira) Original uploader was Nmnogueira at en.wikipedia

File:Fluorescence microscop.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Fluorescence_microscop.jpg *License*: unknown *Contributors*: Masur

File:Microscope And Digital Camera.JPG *Source*: http://en.wikipedia.org/windex.php?title=File:Microscope_And_Digital_Camera.JPG *License*: GNU Free Documentation License *Contributors*: User:Zephyris

File:FluorescenceFilters 2008-09-28.svg *Source*: http://en.wikipedia.org/windex.php?title=File:FluorescenceFilters_2008-09-28.svg *License*: unknown *Contributors*: User:Mastermolch

File:FluorescentCells.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:FluorescentCells.jpg *License*: Public Domain *Contributors*: DO11.10, Emijrp, NEON ja, Origamiemensch, Splette, Tolanor, 5 anonymous edits

File:Yeast membrane proteins.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Yeast_membrane_proteins.jpg *License*: unknown *Contributors*: User:Masur

File:S cerevisiae septins.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:S_cerevisiae_septins.jpg *License*: Public Domain *Contributors*: Spitfire ch, Philippsen Lab, Biozentrum Basel

File:Dividing Cell Fluorescence.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Dividing_Cell_Fluorescence.jpg *License*: unknown *Contributors*: Will-moore-dundee

File:HeLa Hoechst 33258.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:HeLa_Hoechst_33258.jpg *License*: Public Domain *Contributors*: TenOfAllTrades

File:FISH 13 21.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:FISH_13_21.jpg *License*: Public Domain *Contributors*: Gregor1976

File:Bloodcell sun flares pathology.jpeg *Source*: http://en.wikipedia.org/windex.php?title=File:Bloodcell_sun_flares_pathology.jpeg *License*: Public Domain *Contributors*: Birindand, Karelj, NEON ja, 1 anonymous edits

File:Carboxysome 3 images.png *Source*: http://en.wikipedia.org/windex.php?title=File:Carboxysome_3_images.png *License*: Creative Commons Attribution 3.0 *Contributors*: Prof. Todd O. Yeates, UCLA Dept. of Chem. and Biochem.

Image:Abeta-PS3.png *Source*: http://en.wikipedia.org/windex.php?title=File:Abeta-PS3.png *License*: unknown *Contributors*: Gatoatigrado, McLoaf, Million Moments, SirGrant, Skier Dude

Image:FAHMon.png *Source*: http://en.wikipedia.org/windex.php?title=File:FAHMon.png *License*: GNU General Public License *Contributors*: User:SirGrant

Image:GPU - F@H.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:GPU_-_F@H.jpg *License*: unknown *Contributors*: Ofbarea

Image:FAH-tflops.PNG *Source*: http://en.wikipedia.org/windex.php?title=File:FAH-tflops.PNG *License*: Public Domain *Contributors*: Kaleb zero, 5 anonymous edits

Image:LifeWithPlayStation Folding.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:LifeWithPlayStation_Folding.jpg *License*: unknown *Contributors*: ChimpanzeeUK

Image:FAH-SMP.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:FAH-SMP.jpg *License*: unknown *Contributors*: Bovineone, FearTec, SYSS Mouse

Image:Tir parabòlic.png *Source*: http://en.wikipedia.org/windex.php?title=File:Tir_parabòlic.png *License*: GNU Free Documentation License *Contributors*: Abdullah Köroğlu, Iradigalesc, Petri Krohn

Image:physicsdomains.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Physicsdomains.jpg *License*: unknown *Contributors*: User:Loodog

File:Mechanics Overview Table.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Mechanics_Overview_Table.jpg *License*: unknown *Contributors*: User:Saeed.Veradi

Image:Newtons laws in latin.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Newtons_laws_in_latin.jpg *License*: Public Domain *Contributors*: JdH, Man vyi, Tttrung, Wst, 2 anonymous edits

Image:Skaters showing newtons third law.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Skaters_showing_newtons_third_law.svg *License*: Public Domain *Contributors*: Benjamin Crowell (Wikipedia user bcrowell)

Image:Mdalgorithm.PNG *Source*: http://en.wikipedia.org/windex.php?title=File:Mdalgorithm.PNG *License*: Public Domain *Contributors*: User:Knordlun

Image:Electron correlation.png *Source*: http://en.wikipedia.org/windex.php?title=File:Electron_correlation.png *License*: Public Domain *Contributors*: User:Karol Langner

Image:Monte carlo method.svg *Source*: http://en.wikipedia.org/windex.php?title=File:Monte_carlo_method.svg *License*: Public Domain *Contributors*: --pbroks13talk? Original uploader was Pbroks13 at en.wikipedia

Image:Protein folding.png *Source*: http://en.wikipedia.org/windex.php?title=File:Protein_folding.png *License*: Public Domain *Contributors*: DrKjaergaard, PatríciaR

Image:Protein folding schematic.png *Source*: http://en.wikipedia.org/windex.php?title=File:Protein_folding_schematic.png *License*: Public Domain *Contributors*: User:Tomixdf

File:Sarfus.DNABiochip.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Sarfus.DNABiochip.jpg *License*: unknown *Contributors*: Nanolane

File:Rothemund-DNA-SierpinskiGasket.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Rothemund-DNA-SierpinskiGasket.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Antony-22

File:Signal transduction v1.png *Source*: http://en.wikipedia.org/windex.php?title=File:Signal_transduction_v1.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Roadnottaken at en.wikipedia

Image:Holliday Junction.png *Source*: http://en.wikipedia.org/windex.php?title=File:Holliday_Junction.png *License*: Public Domain *Contributors*: Ahruman, Crux, Infrogmation, TimVickers, Wickey

Image:Holliday junction coloured.png *Source*: http://en.wikipedia.org/windex.php?title=File:Holliday_junction_coloured.png *License*: GNU Free Documentation License *Contributors*: Original uploader was Zephyris at en.wikipedia

Image:Mao-DX-schematic.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Mao-DX-schematic.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Antony-22

Image:Mao-DXarray-schematic.gif *Source*: http://en.wikipedia.org/windex.php?title=File:Mao-DXarray-schematic.gif *License*: Creative Commons Attribution 2.5 *Contributors*: Antony-22

Image:SierpinskiTriangle.svg *Source*: http://en.wikipedia.org/windex.php?title=File:SierpinskiTriangle.svg *License*: Public Domain *Contributors*: User:PiAndWhippedCream

Image:Rothemund-DNA-SierpinskiGasket.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Rothemund-DNA-SierpinskiGasket.jpg *License*: Creative Commons Attribution 2.5 *Contributors*: Antony-22

Image:Hydrogen-bonded Self-assembly AngewChemIntEd 1998 v37 p75.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Hydrogen-bonded_Self-assembly_AngewChemIntEd_1998_v37_p75.jpg *License*: GNU Free Documentation License *Contributors*: M stone

Image:DNA nanostructures.png *Source*: http://en.wikipedia.org/windex.php?title=File:DNA_nanostructures.png *License*: unknown *Contributors*: (Images were kindly provided by Thomas H. LaBean and Hao Yan.)

Image:Community of Cells.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:Community_of_Cells.jpg *License*: Public Domain *Contributors*: Mdd

Image:USDAbacteria.jpg *Source*: http://en.wikipedia.org/windex.php?title=File:USDAbacteria.jpg *License*: Public Domain *Contributors*: JWSchmidt, 1 anonymous edits

Image:Notchccr.gif *Source*: http://en.wikipedia.org/windex.php?title=File:Notchccr.gif *License*: Public Domain *Contributors*: JWSchmidt

Image:signal transduction pathways.png *Source*: http://en.wikipedia.org/windex.php?title=File:Signal_transduction_pathways.png *License*: Public Domain *Contributors*: Original uploader was Boghog2 at en.wikipedia

Image:MAPKpathway.png *Source*: http://en.wikipedia.org/windex.php?title=File:MAPKpathway.png *License*: GNU Free Documentation License *Contributors*: JWSchmidt

License

Version 1.2, November 2002

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
14. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright (c) YEAR YOUR NAME.
>
> Permission is granted to copy, distribute and/or modify this document
>
> under the terms of the GNU Free Documentation License, Version 1.2
>
> or any later version published by the Free Software Foundation;
>
> with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
>
> A copy of the license is included in the section entitled "GNU
>
> Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the
>
> Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.